

Dokumentvorlage, Version vom 16.12.2021

# **Dossier zur Nutzenbewertung gemäß § 35a SGB V**

*Luspatercept (Reblozyl®)*

Bristol-Myers Squibb GmbH & Co. KGaA

## **Modul 4 C**

*Erwachsene Patient:innen mit Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist*

Medizinischer Nutzen und  
medizinischer Zusatznutzen,  
Patientengruppen mit therapeutisch  
bedeutsamem Zusatznutzen

Stand: 23.03.2023

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>4</b>
<b>Abbildungsverzeichnis</b> .....	<b>9</b>
<b>Abkürzungsverzeichnis</b> .....	<b>11</b>
<b>4 Modul 4 – allgemeine Informationen</b> .....	<b>14</b>
4.1 Zusammenfassung der Inhalte von Modul 4.....	15
4.2 Methodik .....	27
4.2.1 Fragestellung .....	27
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung .....	28
4.2.3 Informationsbeschaffung .....	30
4.2.3.1 Studien des pharmazeutischen Unternehmers .....	30
4.2.3.2 Bibliografische Literaturrecherche .....	30
4.2.3.3 Suche in Studienregistern/ Studienergebnisdatenbanken .....	31
4.2.3.4 Suche auf der Internetseite des G-BA.....	33
4.2.3.5 Selektion relevanter Studien .....	34
4.2.4 Bewertung der Aussagekraft der Nachweise.....	35
4.2.5 Informationssynthese und -analyse .....	37
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien .....	37
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien.....	38
4.2.5.2.1 Patientencharakteristika .....	38
4.2.5.2.2 Patientenrelevante Endpunkte .....	39
4.2.5.2.3 Verwendete statistische Methodik .....	50
4.2.5.3 Meta-Analysen.....	52
4.2.5.4 Sensitivitätsanalysen .....	53
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren.....	60
4.2.5.6 Indirekte Vergleiche .....	62
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen .....	65
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	65
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel .....	65
4.3.1.1.1 Studien des pharmazeutischen Unternehmers.....	65
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche.....	67
4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken.....	69
4.3.1.1.4 Studien aus der Suche auf der Internetseite des G-BA .....	70
4.3.1.1.5 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel.....	70
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel.....	71
4.3.1.2.1 Studiendesign und Studienpopulationen .....	71
4.3.1.2.2 Verzerrungspotenzial auf Studienebene.....	82

4.3.1.3	Ergebnisse aus randomisierten kontrollierten Studien.....	83
4.3.1.3.1	Endpunkte – RCT.....	84
4.3.1.3.1.1	Mortalität .....	87
4.3.1.3.1.2	Morbidität – Vermeidung von Transfusionen .....	90
4.3.1.3.1.3	Morbidität – Kontrolle der Eisenüberladung.....	106
4.3.1.3.1.4	Morbidität – Symptomatik gemäß NTDT-PRO .....	109
4.3.1.3.1.5	Morbidität – Symptomatik gemäß PGI-S und PGI-C .....	122
4.3.1.3.1.6	Morbidität – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT).....	136
4.3.1.3.1.7	Lebensqualität gemäß SF-36 .....	139
4.3.1.3.1.8	Lebensqualität gemäß FACIT-F.....	152
4.3.1.3.1.9	Sicherheit .....	166
4.3.1.3.2	Subgruppenanalysen – RCT.....	179
4.3.1.4	Liste der eingeschlossenen Studien - RCT .....	206
4.3.2	Weitere Unterlagen.....	206
4.3.2.1	Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien .....	206
4.3.2.1.1	Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche .....	206
4.3.2.1.2	Charakteristika der Studien für indirekte Vergleiche.....	206
4.3.2.1.3	Ergebnisse aus indirekten Vergleichen .....	207
4.3.2.1.3.1	<Endpunkt xxx> – indirekte Vergleiche aus RCT .....	207
4.3.2.1.3.2	Subgruppenanalysen – indirekte Vergleiche aus RCT .....	210
4.3.2.1.4	Liste der eingeschlossenen Studien – indirekte Vergleiche aus RCT... ..	210
4.3.2.2	Nicht randomisierte vergleichende Studien .....	210
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien .....	210
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	211
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien .....	212
4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien....	212
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien .....	213
4.3.2.2.4	Liste der eingeschlossenen Studien – nicht randomisierte vergleichende Studien .....	214
4.3.2.3	Weitere Untersuchungen.....	214
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen .....	214
4.3.2.3.2	Charakteristika der weiteren Untersuchungen .....	215
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen .....	215
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen .....	215
4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen .....	216
4.3.2.3.4	Liste der eingeschlossenen Studien – weitere Untersuchungen.....	216
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	216
4.4.1	Beurteilung der Aussagekraft der Nachweise .....	216
4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	217
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht .....	228
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte .....	228
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	228

4.5.2	Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen.....	228
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen .....	228
4.5.4	Verwendung von Surrogatendpunkten.....	229
4.6	Referenzliste.....	230
<b>Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche .....</b>		<b>235</b>
<b>Anhang 4-B : Suchstrategien – Suche in Studienregistern/ Studienergebnisdatenbanken.....</b>		<b>238</b>
<b>Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche).....</b>		<b>240</b>
<b>Anhang 4-D : Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern/ Studienergebnisdatenbanken).....</b>		<b>241</b>
<b>Anhang 4-E : Methodik der eingeschlossenen Studien – RCT .....</b>		<b>248</b>
<b>Anhang 4-F : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten .....</b>		<b>264</b>
<b>Anhang 4-G : Weitere Analysen .....</b>		<b>287</b>

**Tabellenverzeichnis**

	<b>Seite</b>
Tabelle 4-1: Zusammenfassung der Therapieeffekte von Luspatercept .....	18
Tabelle 4-2: Kriterien für den Ein- und Ausschluss von Studien in die Nutzenbewertung .....	28
Tabelle 4-3: Übersicht der herangezogenen Responseschwellen für den Endpunkt gesundheitsbezogene Lebensqualität gemessen anhand des FACIT-F.....	48
Tabelle 4-4: Subgruppenanalysen der Studie BEYOND .....	61
Tabelle 4-5: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel .....	66
Tabelle 4-6: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	67
Tabelle 4-7: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern / Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel .....	69
Tabelle 4-8: Relevante Studien aus der Suche auf der Internetseite des G-BA – RCT mit dem zu bewertenden Arzneimittel.....	70
Tabelle 4-9: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	71
Tabelle 4-10: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel .....	72
Tabelle 4-11: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	73
Tabelle 4-12: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel.....	73
Tabelle 4-13: Studien- und Therapieabbrüche bis zum primären Datenschnitt.....	81
Tabelle 4-14: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	82
Tabelle 4-15: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel .....	83
Tabelle 4-16: Operationalisierung des Endpunkts „Mortalität“.....	87
Tabelle 4-17: Bewertung des Verzerrungspotenzials für den Endpunkt „Mortalität“ in der Studie BEYOND .....	87
Tabelle 4-18: Ergebnisse für den Endpunkt „Mortalität“ – Anteil der Todesfälle aus der Studie BEYOND (Hauptanalyse).....	89
Tabelle 4-19: Operationalisierung des Endpunkts „Morbidität – Vermeidung von Transfusionen“ .....	90
Tabelle 4-20: Bewertung des Verzerrungspotenzials für den Endpunkt „Morbidität – Vermeidung von Transfusionen“ in der Studie BEYOND .....	92
Tabelle 4-21: Beobachtungszeiten für den Endpunkt „Morbidität – Vermeidung von Transfusionen“ in der Studie BEYOND .....	93

Tabelle 4-22: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Zeit bis zur ersten Transfusion aus der Studie BEYOND (Hauptanalyse)..	94
Tabelle 4-23: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Transfusionsfreiheit über mindestens 24 bzw. 48 Wochen aus der Studie BEYOND (Zusatzanalyse).....	96
Tabelle 4-24: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Anzahl benötigter Transfusionen aus der Studie BEYOND (Zusatzanalyse).....	98
Tabelle 4-25: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Änderung der Hb-Werte gegenüber Baseline aus der Studie BEYOND (Zusatzanalyse).....	100
Tabelle 4-26: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – erythroides Ansprechen über mindestens 24 bzw. 48 Wochen aus der Studie BEYOND (Zusatzanalyse).....	102
Tabelle 4-27: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – längste Dauer des erythroiden Ansprechens aus der Studie BEYOND (Zusatzanalyse).....	103
Tabelle 4-28: Operationalisierung des Endpunkts „Morbidity – Kontrolle der Eisenüberladung“ .....	106
Tabelle 4-29: Bewertung des Verzerrungspotenzials für Endpunkt „Morbidity – Kontrolle der Eisenüberladung“ in der Studie BEYOND.....	107
Tabelle 4-30: Ergebnisse für den Endpunkt „Morbidity – Kontrolle der Eisenüberladung“ – mittlere Änderung des sFerritin-Spiegels gegenüber Baseline aus der Studie BEYOND ..	108
Tabelle 4-31: Operationalisierung des Endpunkts „Morbidity – Symptomatik gemäß NTDT-PRO“ .....	109
Tabelle 4-32: Bewertung des Verzerrungspotenzials für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ in der Studie BEYOND.....	111
Tabelle 4-33: Beobachtungszeiten für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ in der Studie BEYOND .....	112
Tabelle 4-34: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse).....	113
Tabelle 4-35: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse).....	116
Tabelle 4-36: Operationalisierung des Endpunkts „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ .....	122
Tabelle 4-37: Bewertung des Verzerrungspotenzials für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ in der Studie BEYOND .....	124
Tabelle 4-38: Beobachtungszeiten für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ in der Studie BEYOND.....	124

Tabelle 4-39: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse).....	126
Tabelle 4-40: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des PGI-S und PGI-C zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse).....	130
Tabelle 4-41: Operationalisierung des Endpunkts „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“ .....	136
Tabelle 4-42: Bewertung des Verzerrungspotenzials für den „Endpunkt Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“ in der Studie BEYOND .....	137
Tabelle 4-43: Ergebnisse für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)“ – Änderung des 6MWT gegenüber Baseline aus der Studie BEYOND (Hauptanalyse).....	138
Tabelle 4-44: Operationalisierung des Endpunkts „Lebensqualität – SF-36“ .....	139
Tabelle 4-45: Bewertung des Verzerrungspotenzials den Endpunkt „Lebensqualität – SF-36“ in der Studie BEYOND .....	141
Tabelle 4-46: Beobachtungszeiten für den Endpunkt „Lebensqualität – SF-36“ in der Studie BEYOND .....	141
Tabelle 4-47: Ergebnisse für den Endpunkt „Lebensqualität – SF-36“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse) .....	143
Tabelle 4-48: Ergebnisse für den Endpunkt „Lebensqualität – SF-36“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des SF-36 zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse) .....	147
Tabelle 4-49: Operationalisierung des Endpunkts „Lebensqualität – FACIT-F“ .....	152
Tabelle 4-50: Bewertung des Verzerrungspotenzials für den Endpunkt „Lebensqualität – FACIT-F“ in der Studie BEYOND .....	153
Tabelle 4-51: Beobachtungszeiten für den Endpunkt „Lebensqualität – FACIT-F“ in der Studie BEYOND .....	154
Tabelle 4-52: Ergebnisse für den Endpunkt „Lebensqualität – FACIT-F“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse, Gesamtscore).....	155
Tabelle 4-53: Ergebnisse für den Endpunkt „Lebensqualität – FACIT-F“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Zusatzanalyse, Subskalen) .....	157
Tabelle 4-54: Ergebnisse für den Endpunkt „Lebensqualität – FACIT-F“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des FACIT-F Gesamtscores und Subskalen zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse).....	160
Tabelle 4-55: Operationalisierung des Endpunkts „Sicherheit“ .....	166

Tabelle 4-56: Bewertung des Verzerrungspotenzials für den Endpunkt „Sicherheit“ in der Studie BEYOND .....	168
Tabelle 4-57: Ergebnisse für den Endpunkt „Sicherheit“ – Zeit bis zum ersten UE pro Hauptkategorie aus der Studie BEYOND (Hauptanalyse) .....	169
Tabelle 4-58: Ergebnisse für den Endpunkt „Sicherheit“ – Zeit bis zum ersten UE auf Ebene der SOC und PT aus der Studie BEYOND, die bei mindestens 10 % bzw. 5 % der Patient:innen pro Studienarm auftraten und bei denen sich mindestens auf Ebene der PT ein signifikanter Unterschied zwischen den Studienarmen ergab (Hauptanalyse) .....	172
Tabelle 4-59: Ergebnisse für den Endpunkt „Sicherheit“ – Zeit bis zur ersten Hospitalisierung aus der Studie BEYOND (Zusatzanalyse) .....	175
Tabelle 4-60: Ergebnisse für den Endpunkt „Sicherheit“ – Gründe für eine Hospitalisierung aus der Studie BEYOND .....	177
Tabelle 4-61: Matrix der durchgeführten Subgruppenanalysen .....	180
Tabelle 4-62: Ergebnisse des Interaktionstests der Subgruppenanalysen je Endpunkt .....	182
Tabelle 4-63: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ T/W-Gesamtscore – Zeit bis zur ersten bewertungsrelevanten Verbesserung aus der Studie BEYOND .....	187
Tabelle 4-64: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ T/W-Gesamtscore – Zeit bis zur ersten bewertungsrelevanten Verschlechterung aus der Studie BEYOND .....	189
Tabelle 4-65: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ SoB-Gesamtscore – Zeit bis zur ersten bewertungsrelevanten Verschlechterung aus der Studie BEYOND .....	191
Tabelle 4-66: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S“ – Zeit bis zur ersten bewertungsrelevanten Verschlechterung aus der Studie BEYOND .....	193
Tabelle 4-67: Subgruppenanalysen für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)“ – Änderung des 6MWT gegenüber Baseline aus der Studie BEYOND .....	195
Tabelle 4-68: Subgruppenanalysen für den Endpunkt „Sicherheit“ – Zeit bis zum ersten SUE aus der Studie BEYOND .....	197
Tabelle 4-69: Subgruppenanalysen für den Endpunkt „Sicherheit“ – Zeit bis zum ersten schweren UE aus der Studie BEYOND .....	199
Tabelle 4-70: Subgruppenanalysen für den Endpunkt „Sicherheit“ – Zeit bis zum ersten UE nach SOC Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen aus der Studie BEYOND .....	201
Tabelle 4-71: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche .....	207
Tabelle 4-72: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden .....	208
Tabelle 4-73: Operationalisierung von <Endpunkt xxx> .....	208

Tabelle 4-74: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche .....	209
Tabelle 4-75: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche.....	209
Tabelle 4-76: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien .....	211
Tabelle 4-77: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien .....	212
Tabelle 4-78: Operationalisierung von <Endpunkt xxx>.....	212
Tabelle 4-79: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien .....	213
Tabelle 4-80: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen ....	215
Tabelle 4-81: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen.....	215
Tabelle 4-82: Zusammenfassung der Therapieeffekte von Luspatercept .....	219
Tabelle 4-83: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	228
Tabelle 4-84 (Anhang): Liste der im Volltext gesichteten und ausgeschlossenen Dokumente der bibliografischen Literaturrecherche – Suche nach RCT .....	240
Tabelle 4-85 (Anhang): Studiendesign und -methodik für Studie BEYOND .....	249
Tabelle 4-86 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie BEYOND .....	265

**Abbildungsverzeichnis**

	<b>Seite</b>
Abbildung 4-1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel .....	68
Abbildung 4-2: Übersicht des Studiendesigns der Studie BEYOND .....	78
Abbildung 4-3: Kaplan-Meier Plot der Zeit bis zur ersten Transfusion .....	95
Abbildung 4-4: Line-Plot zum mittleren Verlauf der Hb-Werte.....	101
Abbildung 4-5: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung des NTDT-PRO Gesamtscores in der Domäne a) Müdigkeit/Schwäche und b) Kurzatmigkeit .....	114
Abbildung 4-6: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne a) Müdigkeit/Schwäche und b) Kurzatmigkeit .....	115
Abbildung 4-7: Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß NTDT-PRO Gesamtscore in der Domäne a) Müdigkeit/Schwäche und b) Kurzatmigkeit ...	120
Abbildung 4-8: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung a) des PGI-S und b) des PGI-C .....	127
Abbildung 4-9: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung a) des PGI-S und b) des PGI-C .....	128
Abbildung 4-10: Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß PGI-S .....	132
Abbildung 4-11: Bar Chart zur Änderung der Symptomatik gemäß PGI-C für a) Woche 6 bis Woche 66 und b) Woche 72 bis Woche 132 .....	134
Abbildung 4-12: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung a) des SF-36 PCS und b) des SF-36 MCS.....	144
Abbildung 4-13: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung a) des SF-36 PCS und b) des SF-36 MCS .....	145
Abbildung 4-14: Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß a) SF-36 PCS und b) SF-36 MCS .....	150
Abbildung 4-15: Kaplan-Meier Plot a) der Zeit bis zur ersten bewertungsrelevanten Verbesserung des FACIT-F Gesamtscores und b) der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des FACIT-F Gesamtscores .....	156
Abbildung 4-16: Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß des FACIT-F Gesamtscores .....	164
Abbildung 4-17: Kaplan-Meier Plot der Zeit bis zum ersten UE für a) jegliche UE, b) SUE c) schwere UE (CTCAE $\geq 3$ ) und d) zum Therapieabbruch führende UE .....	171
Abbildung 4-18: Kaplan-Meier Plot der Zeit bis zur ersten Hospitalisierung a) unabhängig von der Ursache und b) ohne Berücksichtigung von Hospitalisierungen aufgrund von UE .	176

Abbildung 4-19: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung des NTDT-PRO Gesamtscores in der Domäne T/W für die Subgruppe a) Hb-Wert zu Baseline < 8,5 g/dl und b) Hb-Wert zu Baseline $\geq$ 8,5 g/dl .....	188
Abbildung 4-20: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne T/W für die Subgruppe a) 6MWT zu Baseline $\leq$ 450 m und b) 6MWT zu Baseline > 450 m .....	190
Abbildung 4-21: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne SoB für die Subgruppe a) mit vorhergehender Splenektomie zu Baseline und b) ohne vorhergehende Splenektomie zu Baseline .....	192
Abbildung 4-22: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des PGI-S für die Subgruppe a) FACIT-Fatigue Skala zu Baseline < 43 Punkte und b) FACIT-Fatigue Skala zu Baseline $\geq$ 43 Punkte .....	194
Abbildung 4-23: Kaplan-Meier Plot der Zeit bis zum ersten SUE für die Subgruppe a) mit vorhergehender Splenektomie und b) ohne vorhergehende Splenektomie .....	198
Abbildung 4-24: Kaplan-Meier Plot der Zeit bis zum ersten schweren UE für die Subgruppe a) mit vorhergehender Splenektomie und b) ohne vorhergehende Splenektomie .....	200
Abbildung 4-25: Kaplan-Meier Plot der Zeit bis zum ersten UE der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ für die Subgruppe a) Alter $\leq$ 32 Jahre, b) Alter > 32 Jahre, c) Region Nordamerika/Europa, d) Region Mittlerer Osten und e) Region Asien-Pazifik .....	204
Abbildung 4-26: Patientenfluss in BEYOND bis zum Datenschnitt vom 14.09.2020 .....	263

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
6MWT	6-Minuten-Gehtest
AMIce	Arzneimittelinformationssystem
ANCOVA	<i>Analysis of covariance</i>
BÄK	Bundesärztekammer
BMI	<i>Body-Mass-Index</i>
BSC	<i>Best-Supportive-Care</i>
CONSORT	Consolidated Standards of Reporting Trials
CSR	Studienbericht ( <i>Clinical Study Report</i> )
CTCAE	Common Terminology Criteria for Adverse Events
DIMDI	Deutsches Institut für Medizinische Dokumentation
DMC	<i>Data Monitoring Committee</i>
DNA	Desoxyribonukleinsäure
d. w.	Trockengewicht ( <i>dry weight</i> )
ECOG	<i>Eastern Cooperative Oncology Group</i>
EG	Europäische Gemeinschaft
EK	Erythrozytenkonzentrat
EMA	<i>European Medicines Agency</i>
EMH	extramedulläre Hämatopoese
ERA	Erythrozyten-Reifungs-Aktivator
EU-CTR	<i>EU Clinical Trials Register</i>
FACIT-F	<i>Functional Assessment of Chronic Illness Therapy–Fatigue</i>
G-BA	Gemeinsamer Bundesausschuss
Hb	Hämoglobin
HbE	Hämoglobin E
HBsAg	Hepatitis B-Virus Oberflächenantigen ( <i>hepatitis B virus surface antigen</i> )
HBV	Hepatitis B-Virus
HCV	Hepatitis C-Virus
HR	Hazard Ratio
ICC	<i>intra-class correlation coefficient</i>

<b>Abkürzung</b>	<b>Bedeutung</b>
ICD	<i>International Classification of Disease</i>
ICH	<i>International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use</i>
ICTRP	<i>International Clinical Trials Registry Platform</i>
IQWiG	Institut für Wirtschaftlichkeit im Gesundheitswesen
IRT	<i>Interactive Response Technology</i>
ITT	Intention-to-Treat
KI	Konfidenzintervall
LIC	Lebereisenkonzentration ( <i>liver iron concentration</i> )
Max	Maximum
MCS	psychisches Wohlbefinden ( <i>mental component summary</i> )
MedDRA	<i>Medical Dictionary for Regulatory Activities</i>
Min	Minimum
MMRM	gemischtes Modell für wiederholte Messungen ( <i>mixed-model repeated measure</i> )
MRT	Magnetresonanztomographie
MTC	<i>Mixed Treatment Comparison</i>
MW	Mittelwert
MWD	Mittelwertdifferenz
N.A.	nicht anwendbar, nicht erreicht oder nicht berechenbar
NCI	<i>National Cancer Institute</i>
NTDT	nicht-transfusionsabhängige Beta-Thalassämie
NTDT-PRO	<i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i>
OLP	offene Behandlungshase ( <i>Open-Label-Treatment-Period</i> )
OR	Odds Ratio
PCS	körperliches Wohlbefinden ( <i>physical component summary</i> )
PGI-C	<i>Patient Global Impression of Change</i>
PGI-S	<i>Patient Global Impression of Severity</i>
PT	bevorzugter Begriff ( <i>preferred term</i> )
RCT	<i>Randomized Controlled Trial</i>
RD	Risikodifferenz
RNA	Ribonukleinsäure

<b>Abkürzung</b>	<b>Bedeutung</b>
RR	relatives Risiko
s.c.	subkutan
SD	Standardabweichung ( <i>standard deviation</i> )
SE	Standardfehler ( <i>standard error</i> )
SF-36	<i>Short Form (36) Health Survey</i>
sFerritin	Serum-Ferritin
SGB	Sozialgesetzbuch
SMD	standardisierte Mittelwertdifferenz
SMQ	<i>Standardised MedDRA Query</i>
SOC	Systemorganklasse ( <i>system organ class</i> )
SoB	NTDT-PRO-Gesamtscore in der Domäne Kurzatmigkeit ( <i>shortness of breath</i> )
STE	<i>Surrogate Threshold Effects</i>
STROBE	<i>Strengthening the Reporting of Observational Studies in Epidemiology</i>
SUE	schwerwiegendes UE
T/W	NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche ( <i>tiredness/weakness</i> )
TDT	transfusionsabhängige Beta-Thalassämie
TREND	<i>Transparent Reporting of Evaluations with Non-Randomized Design</i>
UE	unerwünschtes Ereignis
USA	Vereinigte Staaten von Amerika ( <i>United States of America</i> )
VerfO	Verfahrensordnung des Gemeinsamen Bundesausschusses
WHO	<i>World Health Organization</i>
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 4.4)
- ergänzende Informationen zur Begründung der vorgelegten Unterlagen (Abschnitt 4.5)

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

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

#### **4.1 Zusammenfassung der Inhalte von Modul 4**

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

##### **Fragestellung**

Ziel der vorliegenden Nutzenbewertung ist die Bestimmung des Zusatznutzens von Luspatercept (Reblozyl®) bei Erwachsenen für die Behandlung von Anämie, die mit einer nicht-transfusionsabhängigen Beta-Thalassämie verbunden ist<sup>1</sup>. Die Bewertung soll auf Basis von Ergebnissen aus randomisierten kontrollierten Studien (RCT; Evidenzstufe 1b gemäß § 5 Absatz 6 Verfo) hinsichtlich patientenrelevanter Endpunkte aus den Kategorien Mortalität, Morbidität, Lebensqualität und Sicherheit gegenüber der vom G-BA festgelegten zweckmäßigen Vergleichstherapie (zVT), einer bedarfsgerechten Transfusionstherapie mit Erythrozytenkonzentraten (EK) in Kombination mit einer Chelattherapie gemäß der Zulassung, vorzugsweise als Monotherapie, erfolgen.

Luspatercept ist ein Arzneimittel zur Behandlung eines seltenen Leidens gemäß § 35a Absatz 1 Satz 11 SGB V (Orphan Drug). Der Orphan-Drug-Status wurde am 31. Januar 2023 durch den Ausschuss für Arzneimittel für seltene Leiden (COMP) bestätigt. Das vorliegende Dossier bezieht sich auf ein Anwendungsgebiet von Luspatercept, das nach Überschreiten der 30 Millionen Euro Umsatzgrenze zugelassen wurde.

##### **Datenquellen**

Zur Beantwortung der Fragestellung wurde eine systematische Suche nach relevanten Studien in den Datenbanken MEDLINE, EMBASE und Cochrane, sowie den Studienregistern clinicaltrials.gov, *EU Clinical Trials Register* (EU-CTR) und dem Suchportal der *International Clinical Trials Registry Platform* (ICTRP) vorgenommen. Im Rahmen dieser Suche wurde die Zulassungsstudie BEYOND als einzige relevante RCT identifiziert. Als Datenquellen dienen in erster Linie das Protokoll und der Studienbericht der BEYOND sowie für die Nutzenbewertung durchgeführte Zusatzanalysen.

##### **Ein-/Ausschlusskriterien für Studien**

Um geeignete Studien zur Beantwortung der Fragestellung zu identifizieren, wurden Ein- und Ausschlusskriterien für Studien zum medizinischen Nutzen und Zusatznutzen entsprechend dem Anwendungsgebiet von Luspatercept definiert. Eine Übersicht der Ein- und Ausschlusskriterien ist in Abschnitt 4.2.2 dargestellt. Es erfolgte darauf basierend eine systematische Literaturrecherche und Studienregistersuche für Luspatercept, um relevante Studien für die Nutzenbewertung zu identifizieren.

---

<sup>1</sup> Das vorliegende Dossier bezieht sich ausschließlich auf das Anwendungsgebiet der nicht-transfusionsabhängigen Beta-Thalassämie. Der vollständige Wortlaut dieses Anwendungsgebietes lautet: „Reblozyl wird bei Erwachsenen angewendet für die Behandlung von Anämie, die mit transfusionsabhängiger und nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist (siehe Abschnitt 5.1).“

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

Für die vorliegende Nutzenbewertung steht mit der Zulassungsstudie BEYOND eine relevante RCT für einen direkten Vergleich von Luspatercept mit der zVT zur Verfügung.

Die Bewertung der Aussagekraft der Studie erfolgt unter Berücksichtigung des Studiendesigns und des Verzerrungspotenzials der betrachteten Ergebnisse anhand der Kriterien der Modulvorlage. Die Gesamteinschätzung zum Zusatznutzen wird unter Abwägung der Ergebnisse aller patientenrelevanten Endpunkte der Kategorien Mortalität, Morbidität, Lebensqualität und Sicherheit der Studie BEYOND getroffen.

### **Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen**

Die vorliegende Nutzenbewertung erfolgt auf Basis der randomisierten, kontrollierten Doppelblindstudie BEYOND, in der die Wirksamkeit und Sicherheit von Luspatercept im Vergleich zu Placebo, jeweils ergänzt durch eine bestmöglich unterstützende Therapie zur Symptomkontrolle und Verbesserung der Lebensqualität (*Best Supportive Care*; BSC), bei erwachsenen Patient:innen mit Anämie, die mit einer nicht-transfusionsabhängigen Beta-Thalassämie (NTDT) verbunden ist, untersucht wurde.

Die Patient:innen mussten entsprechend den Einschlusskriterien unter einer dokumentierten Beta-Thalassämie oder HbE/Beta-Thalassämie leiden sowie einen mittleren Hämoglobinwert (Hb)  $\leq 10$  g/dl (bezogen auf die letzten 4 Wochen vor der Randomisierung) aufweisen. Darüber hinaus durften Patient:innen in den 24 Wochen vor Randomisierung maximal 5 EK-Einheiten erhalten haben und mussten in den 8 Wochen vor der Randomisierung gänzlich transfusionsfrei gewesen sein.

Insgesamt wurden 145 Patient:innen in die Studie eingeschlossen und im Verhältnis 2:1 auf die beiden Behandlungsarme Luspatercept + BSC (N = 96) – im Folgenden als „Luspatercept“ bezeichnet – und Placebo + BSC (N = 49) – im Folgenden als „Placebo“ bezeichnet – randomisiert.

Luspatercept wurde entsprechend der Fachinformation in einer Dosierung von 0,60 mg/kg bis maximal 1,25 mg/kg Körpergewicht alle 3 Wochen subkutan verabreicht. Die Placebo-Gabe erfolgte analog zur Luspatercept-Gabe, um die Verblindung aufrecht zu erhalten. Unter BSC wurde insbesondere die bedarfsgerechte Transfusionstherapie mit EK in Kombination mit Eisenchelatoren verstanden, die den derzeitigen Behandlungsstandard im vorliegenden Anwendungsgebiet abbildet. Die Studie beinhaltet eine doppelblinde Behandlungsphase, die so lange andauerte, bis alle Patient:innen eine Therapiedauer von mindestens 48 Wochen erreicht oder die Therapie vorzeitig abgebrochen hatten. Nach Abschluss dieser Phase erfolgte die präspezifizierte Analyse und die Studie wurde entblindet. In der nachfolgenden offenen Behandlungsphase wurde die Behandlung der Patient:innen im Placebo-Arm gemäß einer Empfehlung des *Data Monitoring Committees* (DMC) ebenfalls auf Luspatercept umgestellt. Die Auswertungen im vorliegenden Dossier beruhen auf dem Datenschnitt der präspezifizierten Analyse und um-

fassen somit ausschließlich die doppelblinde Phase der Studie, in der alle Patient:innen gemäß der ursprünglichen Randomisierung behandelt wurden.

Neben der Kontrolle der Erkrankung zeigte Luspatercept in der Studie BEYOND eine für die Patient:innen spürbare **Verbesserung der Symptomatik und der gesundheitsbezogenen Lebensqualität**. Die maßgeblichen Ergebnisse zum medizinischen Nutzen und Zusatznutzen sind in der nachfolgenden Tabelle 4-1 zusammengefasst. Aufgrund unterschiedlicher Beobachtungsdauern zwischen den Behandlungsarmen erfolgten die im Dossier dargestellten Hauptanalysen zu Transfusionen, zur Verbesserung/Verschlechterung der Symptomatik und zur Lebensqualität, sowie zum Auftreten von UE anhand von Ereigniszeitanalysen.

Tabelle 4-1: Zusammenfassung der Therapieeffekte von Luspatercept

Studie BEYOND Endpunkt	Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		Ausmaß des Zusatz- nutzens	
	Effektschätzer [95 %-KI]	p-Wert		
<b>Mortalität</b>				
Todesfälle	N.A. (Anteil: 0 % vs. 0 %)	↔	N.A.	Nicht belegt
<b>Morbidität</b>				
<b>Vermeidung von Transfusionen</b> Zeit bis zur ersten Transfusion	HR = 0,23 [0,11; 0,46]	↑	< 0,0001	<b>Beträchtlicher Zusatznutzen</b>
<b>Regulation der Erythropoese (ergänzend dargestellt)</b> Änderung des Hb (g/dl) bis Woche 48	MWD = 1,35 [1,12; 1,57]	↑	< 0,0001	
<b>Kontrolle der Eisenüberladung (ergänzend dargestellt)</b> Änderung des sFerritin (µg/l) bis Woche 48	MWD = 13,50 [-60,96; 87,96]	↔	0,7205	
<b>Symptomatik gemäß NTDT-PRO</b> <i>Gesamtscore in der Domäne Müdigkeit/Schwäche</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 1,03 [0,61; 1,71]	↔	0,8972	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,94 [0,43; 2,04]	↔	0,8670	
<i>Gesamtscore in der Domäne Kurzatmigkeit</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 1,20 [0,67; 2,14]	↔	0,2348	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 1,00 [0,49; 2,01]	↔	0,9714	
<b>Symptomatik gemäß PGI-S und PGI-C</b> <i>PGI-S</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 2,28 [1,15; 4,50]	↑	<b>0,0244</b>	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,71 [0,38; 1,32]	↔	0,2191	
<i>PGI-C</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 6,44 [3,19; 12,97]	↑	< 0,0001	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,51 [0,15; 1,71]	↔	0,2686	
<b>Körperliche Leistungsfähigkeit</b> Änderung im 6-Minuten Gehstest (Meter) zu Woche 48	MWD = 12,44 [-6,72; 31,59]	↔	0,2011	

Studie BEYOND Endpunkt	Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		Ausmaß des Zusatz- nutzens
	Effektschätzer [95 %-KI]	p-Wert	
<b>Gesundheitsbezogene Lebensqualität</b>			
<b>SF-36 – PCS</b>			<b>Beträchtlicher Zusatznutzen</b>
Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 2,79 [1,06; 7,38]	↑ <b>0,0457</b>	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,66 [0,29; 1,50]	↔ 0,3364	
<b>SF-36 – MCS</b>			
Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 1,44 [0,68; 3,08]	↔ 0,1056	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,59 [0,33; 1,07]	↑ <b>0,0303</b>	
<b>FACIT-F – Gesamtscore</b>			
Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 2,41 [0,98; 5,95]	↑ <b>0,0340</b>	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,78 [0,38; 1,59]	↔ 0,5099	
<b>Sicherheit</b>			
<b>UE-Hauptkategorien – Zeit bis zum ersten Ereignis</b>			<b>Nicht belegt</b>
<u>Schwerwiegende</u> UE (SUE)	HR = 0,29 [0,12; 0,69]	↑ <b>0,0032</b>	
<u>Schwere</u> UE (CTCAE ≥ 3)	HR = 1,07 [0,54; 2,14]	↔ 0,8419	
Zum Therapieabbruch führende UE	HR = 0,29 [0,06; 1,34]	↔ 0,0923	
<b>Signifikante Unterschiede bei <u>jeglichen</u> UE auf Ebene der SOC und PT – Zeit bis zum ersten Ereignis</b>			
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen (SOC)	HR = 1,85 [1,19; 2,86]	↓ <b>0,0065</b>	
Knochenschmerzen (PT)	HR = 7,11 [2,18; 23,15]	↓ <b>0,0001</b>	
Tonsillitis (PT)	HR = 0,13 [0,03; 0,64]	↑ <b>0,0033</b>	
Hypertonie (PT)	HR = 9,32 [1,25; 69,66]	↓ <b>0,0079</b>	
Husten (PT)	HR = 7,24 [0,96; 54,68]	↓ <b>0,0248</b>	
Erkrankungen der Haut und des Unterhautgewebes (SOC)	HR = 2,59 [0,99; 6,77]	↓ <b>0,0434</b>	
<b>Signifikante Unterschiede bei <u>schwer- wiegenden/schweren</u> UE auf Ebene der SOC und PT – Zeit bis zum ersten Ereignis</b>			
<u>Schwerwiegende</u> Infektionen und parasitäre Erkrankungen (SOC)	HR = 0,11 [0,02; 0,57]	↑ <b>0,0017</b>	
<u>Schwere</u> Infektionen und parasitäre Erkrankungen (SOC)	HR = 0,05 [0,01; 0,42]	↑ <b>0,0002</b>	

Studie BEYOND Endpunkt	Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		Ausmaß des Zusatz- nutzens
	Effektschätzer [95 %-KI]	p-Wert	
Primärer Datenschnitt: 14.09.2020 ↑ Signifikanter Effekt zugunsten von Luspatercept; ↔ Kein signifikanter Effekt; ↓ Signifikanter Effekt zuungunsten von Luspatercept <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Eine bewertungsrelevante Verbesserung/Verschlechterung entspricht einer Responseschwelle von 15 % der Skalen-spannweite bzw. im Falle des PGI-C den beiden besten/schlechtesten Bewertungen der Veränderung. BSC: <i>Best Supportive Care</i> ; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy–Fatigue</i> ; HR: <i>Hazard Ratio</i> ; MCS: <i>Mental Component Summary</i> ; MWD: <i>Mittelwertdifferenz</i> ; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> ; PCS: <i>Physical Component Summary</i> ; PGI-C/S: <i>Patient Global Impression of Change/Severity</i> ; PT: <i>Preferred Term</i> ; RR: <i>relatives Risiko</i> SF-36: <i>Short Form (36) Health Survey</i> ; SOC: <i>System Organ Class</i> ; SUE: <i>schwerwiegende UE</i> ; UE: <i>unerwünschte Ereignisse</i>			

Bezogen auf die einzelnen Endpunktkategorien ist das Ausmaß des Zusatznutzens für Luspatercept zur Behandlung von erwachsenen Patient:innen mit Anämie, die mit einer nicht-transfusionsabhängigen Beta-Thalassämie verbunden ist, maßgeblich wie folgt begründet:

### Mortalität

In der doppelblinden Studienphase der Studie BEYOND verstarben keine Patient:innen.

In der Endpunktkategorie Mortalität ist ein Zusatznutzen von Luspatercept gegenüber der zVT somit nicht belegt.

### Morbidität

In der Endpunktkategorie Morbidität wurde der Zusatznutzen anhand der patientenrelevanten Endpunkte *Vermeidung von Transfusionen, Veränderung in der Symptomatik* (NTDT-PRO, PGI-S und PGI-C) *und körperliche Belastbarkeit* (6-Minuten-Gehtest) bewertet. Klinisch relevante, therapiesteuernde Parameter der Anämie (Hb-Werte) und Eisenüberladung (sFerritin), die im vorliegenden Anwendungsgebiet einen besonderen Stellenwert haben, werden ergänzend berichtet (siehe Tabelle 4-1).

### *Vermeidung von Transfusionen und Linderung der Anämie*

In der Ereigniszeitanalyse der Zeit bis zur ersten Transfusion ergab sich für den Vergleich von Luspatercept gegenüber Placebo ein HR von 0,23 (95 %-KI [0,11; 0,46];  $p < 0,0001$ ). Dies entspricht einer statistisch signifikanten **Reduktion des Risikos für Transfusionen um 77 %**. Auch in den zusätzlichen Responderanalysen zur Transfusionsfreiheit über 24 und 48 Wochen ergab sich jeweils ein statistisch signifikanter Vorteil für den Luspatercept-Arm gegenüber dem Placebo-Arm.

Die Vermeidung von Transfusionen unter Luspatercept ging einher mit einer Steigerung und langfristigen Stabilisierung der Hb-Werte und somit einer anhaltenden Linderung der Anämie. Im Durchschnitt ergab sich über die Behandlungsphase bis Woche 48 im Luspatercept-Arm ein mittlerer Anstieg der Hb-Werte um 1,44 g/dl während es im Placebo-Arm nur 0,10 g/dl waren (mittlere Differenz: 1,35 [1,12; 1,57];  $p < 0,0001$ ). Die Erhöhung des mittleren Hb-Spiegels im Luspatercept-Arm wurde dazu über den gesamten Behandlungsverlauf bis zum betrachteten Datenschnitt stabil aufrechterhalten.

Die Vermeidung von Transfusionen und Linderung der Anämie stellen eine beträchtliche Verbesserung des therapierelevanten Nutzens dar, die mit einer klinisch relevanten und für die Patient:innen spürbaren Verbesserung ihres Zustands einhergeht:

- Erstens profitieren die Patient:innen langfristig von eingesparten EK-Transfusionen aufgrund der Vermeidung von Transfusionsrisiken, Folgekomplikationen sowie einer weiteren Eisenüberladung.
- Zweitens ist vor dem chronischen Hintergrund der NTDT die Vermeidung von Transfusionen sowie die Verhinderung eines Übergangs in eine transfusionsabhängige Beta-Thalassämie, die mit regelmäßiger Transfusionstherapie und einem erhöhten Behandlungsaufwand, insbesondere einem zeitlichen Aufwand, einhergeht, von großer Bedeutung für die Betroffenen.
- Drittens führt die durch Luspatercept bedingte Steigerung und Stabilisierung der Hb-Werte (Verbesserung der Erythropoese) nicht nur zu einer langanhaltenden Vermeidung von Transfusionen, sondern ermöglicht im Weiteren auch eine Reduktion Anämie-assoziiierter Symptomatik und folglich auch eine spürbare Verbesserung im Befinden der Patient:innen.

Diese therapiebedingten Verbesserungen der Erkrankung spiegeln sich auch klar in der patientenberichteten Veränderung der Symptomatik und der Lebensqualität wider.

### ***Symptomatik gemäß NTDT-PRO***

Im krankheitsspezifischen Fragebogen NTDT-PRO wurden im Hinblick auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung und Verschlechterung (Hauptanalysen) unter Luspatercept jeweils keine signifikanten Vorteile gegenüber Placebo beobachtet. Demgegenüber zeigten sich in den Zusatzanalysen signifikante Vorteile für Luspatercept: So betrug der Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung des NTDT-PRO-Gesamtscores in der Domäne Müdigkeit/Schwäche zu Woche 24 im Luspatercept-Arm 35,5 % gegenüber 17,9 % im Placebo-Arm ( $p = 0,0434$ ). Dieser signifikante Vorteil setzte sich zu Woche 48 mit 41,7 % gegenüber 11,1 % fort ( $p = 0,0317$ ). Darüber hinaus betrug der Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung des NTDT-PRO-Gesamtscores in der Domäne Kurzatmigkeit zu Woche 24 im Luspatercept-Arm 27,6 % gegenüber 10,3 % im Placebo-Arm ( $p = 0,0332$ ). Dieser Vorteil blieb zu Woche 48 mit 35,4 % gegenüber 11,1 % numerisch erhalten ( $p = 0,0594$ ).

Entsprechende Unterschiede sind auch in den Abbildungen der mittleren Verläufe und den MMRM-Analysen ersichtlich. Zudem lassen die Abbildungen vermuten, dass die Vorteile im Rahmen der Ereigniszeitanalysen aufgrund von vorübergehenden Anfangsverbesserungen im Placebo-Arm ggf. unzureichend abgebildet werden.

Insgesamt sind die Ergebnisse zum NTDT-PRO positiv zu bewerten. Die Zusatzanalysen deuten auf eine spürbare Verbesserung der krankheitsspezifischen Symptomatik unter Luspatercept hin.

### ***Symptomatik gemäß PGI-S und PGI-C***

Für die patientenberichtete Symptomatik gemäß PGI-S und PGI-C zeigte sich in Bezug auf die Zeit bis zur ersten Verbesserung jeweils ein sehr deutlicher, statistisch signifikanter Vorteil für Luspatercept im Vergleich zu Placebo.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des PGI-S ergab sich ein signifikanter Effekt mit einem HR von 2,28 (95 %-KI [1,15; 4,50];  $p = 0,0244$ ), d. h. die **Chance auf eine bewertungsrelevante Verbesserung des PGI-S war unter Luspatercept mehr als doppelt so hoch** wie unter Placebo.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des PGI-C ergab sich ebenfalls ein signifikanter Effekt mit einem HR von 6,44 (95 %-KI [3,19; 12,97];  $p < 0,0001$ ), d. h. die **Chance auf eine bewertungsrelevante Verbesserung des PGI-C war unter Luspatercept mehr als 6-mal höher** als unter Placebo.

Diese Ergebnisse wurden durch die Zusatzanalysen zum Anteil der Patient:innen mit einer signifikanten Verbesserung zu Woche 24 und 48 bestätigt und spiegeln sich darüber hinaus auch eindeutig in den Ergebnissen der MMRM-Analysen und den Verlaufskurven zum PGI-S bzw. in der Abbildung zu den PGI-C-Bewertungen im Studienverlauf wider.

Die Ergebnisse zum PGI-S und PGI-C verdeutlichen, dass die anhaltende Linderung der Anämie entsprechend der Steigerung und langfristigen Stabilisierung der Hb-Werte durch Luspatercept auch zu einer sehr deutlichen und patientenrelevanten Reduktion der durch die Patient:innen wahrgenommenen Symptombelastung führt.

### ***Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest***

In Bezug auf die körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest zeigte sich in der Studie BEYOND kein Unterschied zwischen den Behandlungsarmen.

### ***Fazit Morbidität***

Zusammengenommen führt die Behandlung mit Luspatercept im Vergleich zur zVT zu bedeutenden Vorteilen in Bezug auf die langfristige Vermeidung von Transfusionen in Verbindung mit einer anhaltenden Linderung der Anämie sowie einer spürbaren und deutlichen Reduktion der Krankheitssymptomatik. **Für die Endpunktkategorie Morbidität besteht daher ein beträchtlicher Zusatznutzen von Luspatercept gegenüber der zVT.**

## Lebensqualität

In der Endpunktkategorie gesundheitsbezogene Lebensqualität wurde der Zusatznutzen in der Studie BEYOND anhand eines generischen Fragebogens (SF-36) und eines Fatigue-spezifischen Fragebogens (FACIT-F) bewertet. Die durch die ineffektive Erythropoese bedingte Anämie kann die Lebensqualität von Beta-Thalassämie-Patient:innen maßgeblich beeinträchtigen. Die Aufrechterhaltung oder Verbesserung der Lebensqualität ist im vorliegenden Anwendungsgebiet daher ein wichtiges Therapieziel.

### *SF-36*

Für den Endpunkt SF-36 zeigten sich in der Studie BEYOND im Rahmen der Ereigniszeitanalysen statistisch signifikante Effekte zugunsten von Luspatercept in Bezug auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung des PCS (körperliches Wohlbefinden) und die Zeit bis zur ersten bewertungsrelevanten Verschlechterung des MCS (psychisches Wohlbefinden).

Für den PCS ergab sich bei der Zeit bis zur ersten bewertungsrelevanten Verbesserung ein HR von 2,79 (95 %-KI [1,06; 7,38];  $p = 0,0457$ ), d. h. die **Chance auf eine Verbesserung des körperlichen Wohlbefindens war unter Luspatercept mehr als doppelt so hoch** wie unter Placebo.

Für den MCS ergab sich bei der Zeit bis zur ersten bewertungsrelevanten Verschlechterung ein HR von 0,59 (95 %-KI [0,33; 1,07];  $p = 0,0303$ ), d. h. **das Risiko für eine Verschlechterung des psychischen Wohlbefindens war unter Luspatercept 41 % geringer** als unter Placebo.

### *FACIT-F*

Für den Endpunkt FACIT-F zeigte sich in der Studie BEYOND bei der Zeit bis zur Verbesserung des Gesamtscores ein statistisch signifikanter Effekt zugunsten von Luspatercept.

Für den FACIT-F Gesamtscore ergab sich bei der Zeit bis zur ersten bewertungsrelevanten Verbesserung ein HR von 2,41 (95 %-KI [0,98; 5,95];  $p = 0,0340$ ), d. h. die **Chance auf eine Verbesserung des Gesamtscores der Fatigue-spezifischen Lebensqualität war unter Luspatercept mehr als doppelt so hoch** wie unter Placebo.

### *Fazit Lebensqualität*

Zusammengenommen geht aus den Ergebnissen hervor, dass die anhaltende Linderung der Anämie unter Luspatercept auch zu einer spürbaren Verbesserung der Lebensqualität der NTDT-Patient:innen führt. **Für die Endpunktkategorie gesundheitsbezogene Lebensqualität besteht daher ein beträchtlicher Zusatznutzen von Luspatercept gegenüber der zVT.**

## **Sicherheit**

Die Auswertung der unerwünschten Ereignisse (UE) in der Studie BEYOND zeigt, dass die Behandlung mit Luspatercept im Vergleich zu Placebo insgesamt mit einer signifikanten Reduktion des Risikos für schwerwiegende UE (SUE) sowie jeweils für schwere und schwerwiegende UE der SOC „Infektionen und parasitäre Erkrankungen“ einherging.

Darüber hinaus ergaben sich auf SOC- und PT-Ebene keine weiteren Unterschiede bei schwerwiegenden und schweren UE. Auch bei zum Therapieabbruch führenden UE wurden keine Unterschiede festgestellt.

Bei Betrachtung jeglicher UE auf Ebene der SOC und PT zeigte sich ein Vorteil bei dem PT „Tonsillitis“, während bei der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“, dem PT „Knochenschmerzen“, dem PT „Hypertonie“, dem PT „Husten“ und der SOC „Erkrankungen der Haut und des Unterhautgewebes“ signifikante Unterschiede zuungunsten von Luspatercept festgestellt wurden.

Knochenschmerzen und Hypertonie sind für Luspatercept bekannte unerwünschte Arzneimittelwirkungen und entsprechend in der Fachinformation von Reblozyl<sup>®</sup> aufgeführt. Das häufigere Auftreten von Hypertonie unter Luspatercept wird in den ersten 8 bis 12 Monaten beobachtet und kann eine Anpassung oder Verzögerung der Luspatercept-Dosis oder eine Behandlung gemäß den Empfehlungen der aktuellen klinischen Leitlinien für Hypertonie erfordern. Das häufigere Auftreten von Knochenschmerzen unter Luspatercept ist auf die ersten 3 Monate begrenzt. Die meisten entsprechenden UE unter Luspatercept (32 von 35 Ereignissen) waren von leichter bis moderater Schwere (Grad 1 oder 2) und nur 3 UE wurden als „schwer“ eingestuft (Grad 3). In der klinischen Praxis sind die anfänglichen Knochenschmerzen, sofern erforderlich, mit Analgetika gut therapierbar. Kein Patient und keine Patientin brach die Behandlung mit Luspatercept aufgrund von Knochenschmerzen ab. Das zeitlich begrenzte Auftreten von Knochenschmerzen deutet auf eine verstärkte Knochenmarksaktivität (Stimulation der Erythropoese), d. h. auf die gewünschte Wirkung von Luspatercept, hin.

### ***Fazit Sicherheit***

Insgesamt besteht im Hinblick auf UE ein mindestens ausgewogenes Bild für Luspatercept im Vergleich zu Placebo, das sich weiterhin mit dem bekannten, gut tolerierbaren Sicherheits- und Verträglichkeitsprofil von Luspatercept deckt. Aufgrund des höheren Risikos für spezifische Ereignisse (insbesondere Knochenschmerzen und Hypertonie) wird trotz der signifikanten Reduktion der Gesamtrate schwerwiegender UE für die Endpunktkategorie Sicherheit kein Zusatznutzen von Luspatercept gegenüber der zVT abgeleitet.

## **Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen**

Die Beta-Thalassämie ist eine seltene, erblich bedingte, hämatologische Erkrankung mit einer erheblichen Krankheitslast und Beeinträchtigung der Lebensqualität. Dabei spielen sowohl erkrankungsbedingte als auch therapiebedingte (z. B. transfusionsbedingte klinische und psychosoziale) Komplikationen eine große Rolle. Im Zentrum der Komplikationen steht die

Anämie infolge der ineffektiven Erythropoese aufgrund des vorliegenden Erythrozyten-Reifungs-Defekts.

Für Patient:innen im Anwendungsgebiet von Luspatercept gibt es einen großen therapeutischen Bedarf nach einem Arzneimittel, das den Erythrozyten-Reifungs-Defekt adressiert und mit einer Erhöhung und Stabilisierung des Hb-Wertes die Anämie lindert. NTDT-Patient:innen leiden häufig über ausgedehnte Zeiträume unter niedrigen Hb-Werten sowie entsprechenden Anämie-assoziierten Symptomen und erfahren nur selten eine Linderung. Zudem treten Folgekomplikationen, bedingt durch die Grunderkrankung, häufiger und mitunter schwerwiegender bei NTDT-Patient:innen auf, da diese im Gegensatz zu den TDT-Patient:innen keine kontinuierliche Transfusionstherapie zur Unterdrückung der eigenen ineffektiven Erythropoese erhalten. Aber auch die Entwicklung einer regelmäßigen Transfusionsnotwendigkeit und damit ein Wechsel des klinischen Bildes hin zu dem einer TDT ist möglich.

Luspatercept ist der erste Vertreter der neuen Wirkstoffklasse der **Erythrozyten-Reifungs-Aktivatoren** (ERA) und damit die erste aktive, medikamentöse Therapie zur Behandlung der Anämie bei Erwachsenen mit nicht-transfusionsabhängiger Beta-Thalassämie. Die Therapie mit Luspatercept adressiert gezielt den der Anämie zugrundeliegenden Erythrozyten-Reifungs-Defekt. Dies führt zu einer Steigerung und langfristigen Stabilisierung der Hb-Werte, wodurch einerseits Transfusionen und potenzielle Folgekomplikationen vermieden werden können und andererseits Symptomatik, Lebensqualität und Autonomie der Patient:innen verbessert werden. Damit deckt Luspatercept den großen therapeutischen Bedarf im vorliegenden Anwendungsgebiet.

Die Ergebnisse zum medizinischen Nutzen und Zusatznutzen von Luspatercept beruhen auf der doppelblinden, randomisierten, kontrollierten Zulassungsstudie BEYOND. Gemäß § 5 Absatz 6 des 5. Kapitels der Verfo entspricht die Studie BEYOND als RCT der Evidenzstufe 1b. Somit ist die Studie auf Basis signifikanter Ergebnisse bei Endpunkten mit einem niedrigen Verzerrungspotenzial geeignet, um eine Zusatznutzenaussage auf Ebene eines Hinweises abzuleiten. Im vorliegenden Dossier wurde zum Zwecke einer konservativen Einschätzung für unerwünschte Ereignisse sowie für die patientenberichteten Endpunkte zur Lebensqualität und zur Symptomatik gemäß NTDT-PRO ein erhöhtes Verzerrungspotenzial abgeleitet. Demgegenüber liegt allerdings u. a. für die Morbiditätsendpunkte „Vermeidung von Transfusionen“ und „Symptomatik gemäß PGI-S und PGI-C“ jeweils ein niedriges Verzerrungspotenzial vor. Folglich können Aussagen zum Zusatznutzen auf Basis der signifikanten Ergebnisse bei diesen Endpunkten als „Hinweis“ eingestuft werden.

Für die Gesamtschau des Zusatznutzens sind folgende Ergebnisse relevant:

- **Morbidität:**
  - **Beträchtlicher Zusatznutzen** in Hinblick auf die **Vermeidung von Transfusionen**. Die signifikante Reduktion des Risikos für Transfusionen bzw. der Erhalt der Transfusionsfreiheit unter Luspatercept ging in der Studie BEYOND gleichzeitig mit

einer deutlichen Steigerung und langfristigen Stabilisierung der Hb-Werte und somit einer anhaltenden Linderung der Anämie einher. Diese Ergebnisse zeigen, dass die Verbesserung der Erythropoese durch Luspatercept den Patient:innen einerseits den langfristigen Erhalt einer Transfusionsunabhängigkeit – und, damit einhergehend, eine Vermeidung von Transfusionsrisiken, Folgekomplikationen sowie weiteren Belastungen – bietet und ihnen andererseits eine Reduktion der Anämie-assoziierten Symptomatik und Verbesserung ihres Befindens ermöglichen kann.

- **Beträchtlicher Zusatznutzen** hinsichtlich der **Symptomatik gemäß PGI-S und PGI-C**. Als selbstberichtete Fragebögen spiegeln die Endpunkte die unmittelbare Erfahrung und Wahrnehmung der Patient:innen wider. Die Ergebnisse zeigen, dass die Therapie mit Luspatercept gegenüber der zVT zu einer sehr deutlichen und patientenrelevanten Reduktion der durch die Patient:innen wahrgenommenen Symptombelastung führt.
- **Lebensqualität:**
  - **Beträchtlicher Zusatznutzen** in Bezug auf das **körperliche und psychische Wohlbefinden (SF-36) sowie den Gesamtscore der Fatigue-spezifischen Lebensqualität (FACIT-F)**. Die Ergebnisse zeigen, dass die Behandlung mit Luspatercept gegenüber der zVT mit einer spürbaren Verbesserung in verschiedenen Bereichen der Lebensqualität verbunden ist. Aufgrund der chronischen Natur der Erkrankung ist die gesundheitsbezogene Lebensqualität bei Patient:innen mit NTDT von besonderer Relevanz.

In der Gesamtschau zeigt sich durch Luspatercept eine beträchtliche Verbesserung des therapie-relevanten Nutzens, begründet in signifikanten und spürbaren Behandlungsunterschieden, die einer deutlichen Verbesserung der Morbidität und gesundheitsbezogenen Lebensqualität der Patient:innen entsprechen.

**Insgesamt** ergibt sich für die Behandlung von erwachsenen Patient:innen mit Anämie, die mit einer nicht-transfusionsabhängigen Beta-Thalassämie verbunden ist, auf Basis der Ergebnisse der Studie BEYOND ein **Hinweis auf einen beträchtlichen Zusatznutzen von Luspatercept** gegenüber der 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.*

Ziel der vorliegenden Nutzenbewertung ist die Bestimmung des Zusatznutzens von Luspatercept (Reblozyl®) bei Erwachsenen für die Behandlung von Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist. Die Bewertung erfolgt auf Basis von Ergebnissen aus randomisierten kontrollierten Studien (RCT; Evidenzstufe 1b gemäß § 5 Absatz 6 Verfo) hinsichtlich patientenrelevanter Endpunkte aus den Kategorien Mortalität, Morbidität, Lebensqualität und Sicherheit gegenüber der vom G-BA festgelegten zweckmäßigen Vergleichstherapie (zVT), einer bedarfsgerechten Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Chelattherapie gemäß der Zulassung, vorzugsweise als Monotherapie.

Luspatercept ist ein Arzneimittel zur Behandlung eines seltenen Leidens gemäß § 35a Absatz 1 Satz 11 SGB V (Orphan Drug). Der Orphan-Drug-Status wurde am 31. Januar 2023 durch den Ausschuss für Arzneimittel für seltene Leiden (COMP) bestätigt [1]. Das vorliegende Dossier bezieht sich auf ein Anwendungsgebiet von Luspatercept, das nach Überschreiten der 30 Millionen Euro Umsatzgrenze zugelassen wurde.

#### 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.*

Um relevante Studien für die Beantwortung der in Abschnitt 4.2.1 formulierten Fragestellung zu identifizieren, wurden Ein- und Ausschlusskriterien hinsichtlich der Patientenpopulation, der Intervention, der erhobenen Endpunkte und des Studiendesigns definiert. Eine Übersicht der Ein- und Ausschlusskriterien findet sich in Tabelle 4-2.

Studien, die die genannten Kriterien nicht erfüllten, wurden ausgeschlossen. Studien, die die Kriterien nur teilweise erfüllten, wurden dann eingeschlossen, wenn mindestens 80 % der Studienpopulation die Kriterien erfüllte oder eine den Kriterien entsprechende Auswertung von Teilpopulation(en) möglich war [2].

Tabelle 4-2: Kriterien für den Ein- und Ausschluss von Studien in die Nutzenbewertung

Selektionskriterien	Einschlusskriterien	Ausschlusskriterien
1 Patientenpopulation	E1 Erwachsene Patient:innen mit Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist	A1 Population nicht wie in E1 definiert (d. h. Patientenpopulation, die nicht der Zulassung von Luspatercept entspricht)
2 Intervention	E2 Luspatercept gemäß Fachinformation	A2 Intervention nicht wie in E2 definiert
3 Vergleichstherapie	E3 Bedarfsgerechte Transfusions-therapie mit Erythrozytenkonzentraten in Kombination mit einer Chelattherapie gemäß der Zulassung, vorzugsweise als Monotherapie	A3 Die Vergleichstherapie entspricht nicht der zVT

Selektionskriterien	Einschlusskriterien	Ausschlusskriterien
4 Endpunkte	E4 Mindestens ein patientenrelevanter Endpunkt (Mortalität, Morbidität, Lebensqualität, Sicherheit) wurde berichtet (siehe auch Abschnitt 4.2.5.2.2)	A4 Keine patientenrelevanten Endpunkte berichtet
5 Studiendesign	E5 Studie im RCT-Design oder Meta-Analyse aus RCT	A5 Keine Studie im RCT-Design oder daraus abgeleitete Meta-Analyse
6 Studienlänge	E6 Studiendauer $\geq$ 24 Wochen	A6 Studiendauer $<$ 24 Wochen
7 Ergebnisdarstellung	E7 Berichterstattung liefert ausreichende Informationen zur Beurteilung von Methodik und Ergebnissen in Anlehnung an die CONSORT-Kriterien (z. B. Studienbericht, Volltextpublikation in Fachzeitschriften oder ausführliche Ergebnisdarstellung in Studienregister verfügbar); Ergebnisse stammen von einzelnen Studien oder Meta-Analysen auf Grundlage der Studienrohdaten; Publikation in Deutsch oder Englisch	A7 Berichterstattung liefert keine ausreichenden Informationen zur Beurteilung von Methodik und Ergebnissen (z. B. Konferenzbeitrag); Meta-Analysen auf Grundlage bereits publizierter Studiendaten (Sekundärliteratur); Publikation in anderer Sprache als Deutsch oder Englisch
8 Registereintrag	E8 -	A8 Treffer ist Eintrag aus anderweitig durchsuchtem Register (d. h. clinicaltrials.gov, EU-CTR oder ICTRP) [nur anwendbar auf Treffer aus der Suche in der Literaturdatenbank der Cochrane Library]

CONSORT: *Consolidated Standards of Reporting Trials*; EU-CTR: *EU Clinical Trials Register*; ICTRP: *International Clinical Trials Registry Platform*; RCT: randomisierte kontrollierte Studie (*randomized controlled trial*); zVT: zweckmäßige Vergleichstherapie

Die Fragestellung der vorliegenden Nutzenbewertung bezieht sich auf das gemäß Zulassung definierte Anwendungsgebiet von Luspatercept bei Erwachsenen für die Behandlung von Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist. Die Selektionskriterien wurden dementsprechend gewählt.

Das Selektionskriterium 1 (Patientenpopulation) definiert die relevante Patientenpopulation gemäß der Zulassung von Luspatercept in dem vorliegendem Anwendungsgebiet [3]. Die zu untersuchende Intervention (Selektionskriterium 2) ist Luspatercept in der zugelassenen Darreichungsform und Dosierung [3]. Die erforderliche Vergleichstherapie (Selektionskriterium 3) entspricht der zVT gemäß der Festlegung durch den G-BA, dem Zulassungsstatus und der praktischen Anwendung (detaillierte Begründung siehe Modul 3 C Kapitel 3.1). Um für die Nutzenbewertung von Relevanz zu sein, müssen Studien mindestens einen patientenrelevanten Endpunkt aus einer der Kategorien Mortalität, Morbidität, Lebensqualität oder

Sicherheit berichten (Selektionskriterium 4; vgl. zu patientenrelevanten Endpunkten auch Abschnitt 4.2.5.2). Um eine hohe Aussagekraft der Ergebnisse sicherzustellen, wurden ausschließlich Studien im RCT-Design (Evidenzgrad 1b) oder Meta-Analysen aus RCT eingeschlossen (Selektionskriterium 5). Um relevante Behandlungseffekte im Kontext der zugrundeliegenden chronischen Erkrankung nachweisen zu können, ist eine Studiendauer von mindestens 24 Wochen erforderlich (Selektionskriterium 6). Das Selektionskriterium 7 (Ergebnisdarstellung) beinhaltet die Einschränkung auf Studien, zu denen ausreichende Informationen verfügbar sind, um eine adäquate Bewertung und Einordnung der Ergebnisse sicherzustellen. Das Selektionskriterium 8 (Registereintrag) betrifft ausschließlich Treffer aus anderweitig durchsuchten Registern, die in einer separaten Suche in Studienregistern identifiziert werden (anwendbar auf Treffer aus der Suche in der Literaturdatenbank der Cochrane Library).

### 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 bibliografische Literaturrecherche für das zu bewertende Arzneimittel Luspatercept wurde am 02.02.2023 in den Datenbanken MEDLINE (inklusive „Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily“), EMBASE und Cochrane (*Cochrane Central Register of Controlled Trials*) entsprechend der in Abschnitt 4.2.1 formulierten Fragestellung und den in Abschnitt 4.2.2 definierten Ein- und Ausschlusskriterien durchgeführt. Die Suche wurde in jeder Datenbank einzeln und mit einer jeweils adaptierten Suchstrategie durchgeführt, welche im Anhang 4-A detailliert dokumentiert ist. In MEDLINE und EMBASE wurde die Suche mittels eines validierten Suchfilters auf RCT eingeschränkt. Für die Suche in der Cochrane Datenbank wurde der Suchblock nach der Intervention auf die Kategorie „Trials“ eingeschränkt. Die Suchstrategie wurde weder zeitlich noch sprachlich eingeschränkt.

#### **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://clinicaltrials.gov) ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)), EU Clinical Trials Register (EU-CTR, [www.clinicaltrialsregister.eu](http://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 Studien mit dem zu bewertenden Arzneimittel Luspatercept in Studienregistern wurde am 02.02.2023 in den Studienregistern [clinicaltrials.gov](http://clinicaltrials.gov), *EU Clinical Trials Register* (EU-CTR) und dem Suchportal der *International Clinical Trials Registry Platform* (ICTRP) entsprechend der in Abschnitt 4.2.1 definierten Fragestellung und den Ein- und Ausschlusskriterien gemäß Abschnitt 4.2.2 durchgeführt. Die Suche wurde in jedem Studienregister einzeln und mit einer jeweils adaptierten Suchstrategie durchgeführt.

Um die Vollständigkeit der Ergebnisdarstellung der identifizierten relevanten Studien zu überprüfen, wurde anschließend in den Studienergebnisdatenbanken der *European Medicines Agency* (EMA) (*Clinical Data*) und des Arzneimittelinformationssystems (AMIce) nach den aus der bibliografischen Literaturrecherche und der Recherche in Studienregistern identifizierten, relevanten Studien gezielt gesucht. Als Suchbegriffe wurden verschiedene Varianten der Intervention und Studienidentifikationsnummern verwendet.

Die Recherchestrategie wurde weder zeitlich, sprachlich noch anhand des Status der Studie eingeschränkt. Eine detaillierte Dokumentation der Suchen in den Studienregistern *clinicaltrials.gov*, EU-CTR und dem Suchportal ICTRP befindet sich im Anhang 4-B.

#### 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 –ergebnissen<sup>2</sup>. 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 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.*

---

<sup>2</sup> Köhler M, Haag S, Biester K, Brockhaus AC, McGauran N, Grouven U, Kölsch H, Seay U, Hörn H, Moritz G, Staeck 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

*Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.*

Die Internetseite des G-BA wurde am 06.02.2023 nach der Intervention („Luspatercept“), Indikation („ $\beta$ -Thalassämie“, „beta-Thalassämie“ und „beta Thalassämie“) sowie den Studiennummern („BEYOND“, „NCT03342404“, „ACE-536-B-THAL-002“), die im Rahmen der bibliografischen Literaturrecherche und/oder der Suche in Studienregistern und Studienergebnisdatenbanken als relevant identifiziert wurden, durchsucht.

#### **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.*

Die Selektion relevanter Studien aus den Ergebnissen der bibliografischen Literaturrecherche sowie der Suche in Studienregistern wurde anhand der folgenden Schritte nacheinander von zwei Personen unabhängig voneinander durchgeführt. Sofern sich bezüglich des Ein- und Ausschlusses von Treffern abweichende Beurteilungen ergaben, wurden diese im abschließenden Konsensverfahren durch die beiden beteiligten Personen geklärt.

Als relevant identifizierte Studien wurden durch gezielte Suchen in den Studienergebnisdatenbanken (EMA *Clinical Data* und AMIce) sowie auf der Internetseite des G-BA auf Vollständigkeit hinsichtlich der Verfügbarkeit relevanter Informationen und Studienergebnisse überprüft.

#### **Bibliografische Literaturrecherche**

Die Selektion relevanter Studien aus den Ergebnissen der bibliografischen Literaturrecherche erfolgte in einem vierstufigen Prozess:

1. Entfernung von Duplikaten
2. Ausschluss von Referenzen, bei denen klar erkennbar war, dass es sich um keine Berichtserstattungen handelt, welche ausreichende Informationen zur Beurteilung der Methodik und Ergebnissen liefern
3. Ausschluss nicht relevanter Ergebnisse durch Titel- und Abstract-Screening anhand der in Abschnitt 4.2.2 definierten Ein- und Ausschlusskriterien; Ergebnisse, die nicht eindeutig anhand des Titels und/oder Abstracts ausgeschlossen werden konnten, wurden im 4. Schritt überprüft; ausgeschlossene Meta-Analysen wurden zusätzlich auf die Darstellung relevanter Studien geprüft
4. Ausschluss nicht relevanter Ergebnisse nach Volltext-Screening

## Suche in Studienregistern

Zur Selektion relevanter Studien aus den Ergebnissen der Suche in Studienregistern wurden in einem ersten Schritt solche Studienregistereinträge ausgeschlossen, die den in Abschnitt 4.2.2 festgelegten Einschlusskriterien nicht entsprachen. Grundlage dieser Einschätzung waren alle in den Studienregistern hinterlegten Informationen. Die Selektion relevanter Studien erfolgte für jedes Studienregister einzeln.

### 4.2.4 Bewertung der Aussagekraft der Nachweise

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

#### A: Verzerrungsaspekte der Ergebnisse auf Studienebene

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

#### B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene

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

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

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

Für die Bewertung eines Endpunkts soll für randomisierte Studien zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter A aufgeführten Aspekte als „niedrig“ oder „hoch“ eingestuft werden. Falls diese Einstufung als „hoch“ erfolgt, soll das

Verzerrungspotenzial für den Endpunkt in der Regel auch als „hoch“ bewertet werden, Abweichungen hiervon sind zu begründen. Ansonsten sollen die unter B genannten endpunktspezifischen Aspekte Berücksichtigung finden.

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

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

Die eingeschlossenen, relevanten Studien wurden hinsichtlich ihrer Aussagekraft anhand der zur Verfügung stehenden Informationen (Studienberichte und Publikationen) bewertet.

### **Bewertung des Verzerrungspotenzials**

Die Bewertung des Verzerrungspotenzials erfolgte sowohl auf Studienebene als auch auf Endpunktebene mithilfe der Bewertungsbögen in Anhang 4-F gemäß den obenstehenden Vorgaben des G-BA.

#### ***Verzerrungsaspekte der Ergebnisse auf Studienebene***

In einem ersten Schritt wurde das endpunktübergreifende Verzerrungspotenzial anhand der folgenden Komponenten als „hoch“ oder „niedrig“ bewertet:

- Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Verblindung der Patient:innen sowie der behandelnden Personen
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

#### ***Verzerrungsaspekte der Ergebnisse auf Endpunktebene***

Bei einem hohen Verzerrungspotenzial auf Studienebene ist davon auszugehen, dass auch die Endpunkte hoch verzerrt sind, sodass auch hier eine Einstufung des Verzerrungspotenzials als „hoch“ vorgenommen wurde. (In Ausnahmen kann das Verzerrungspotenzial auf Endpunktebene aber auch trotz hoher Verzerrung auf Studienebene „niedrig“ sein; solche Fälle bedürfen einer entsprechenden Begründung.)

Bei einem niedrigen Verzerrungspotenzial auf Studienebene wurde für jeden Endpunkt einzeln das Verzerrungspotenzial auf Endpunktebene gemäß den folgenden Aspekten geprüft:

- Verblindung der Endpunkterheber:innen
- Umsetzung des ITT-Prinzips
- Ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

Die Einstufungen der Verzerrungspotenziale auf Studien- und Endpunktebene dienen der Diskussion heterogener Studienergebnisse und der Bewertung von deren Aussagesicherheit. Ein Ausschluss von Ergebnissen ausschließlich aufgrund eines hohen Verzerrungspotenzials fand nicht statt.

Eine detaillierte Bewertung der Verzerrungspotenziale ist im Anhang 4-F 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)<sup>3</sup>. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-<sup>4</sup> bzw. STROBE-Statements<sup>5</sup> folgen. Design und Methodik weiterer Untersuchungen sollen gemäß den verfügbaren Standards dargestellt werden.

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

Die Darstellung des Designs und der Methodik der eingeschlossenen Studie erfolgte anhand der Empfehlungen des CONSORT-Statements 2010 (Items 2b bis 14), welches Angaben zur Methodik (Studienhypothesen, Ein- und Ausschlusskriterien, Interventionen, Endpunkte, Gruppenteilung, Verblindung und statistische Methodik) und dem Patientenfluss enthält.

---

<sup>3</sup> Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.

<sup>4</sup> 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.

<sup>5</sup> 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.

Eine detaillierte Beschreibung inklusive eines Flow-Charts, das den Patientenfluss veranschaulicht, befindet sich im Anhang 4-E.

#### 4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien

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

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

*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.*

Für die vorliegende Nutzenbewertung steht mit der Zulassungsstudie BEYOND eine relevante RCT für einen direkten Vergleich von Luspatercept mit der zVT zur Verfügung (vgl. Abschnitt 4.3.1.1).

##### 4.2.5.2.1 Patientencharakteristika

In der Studie BEYOND wurde Luspatercept im Vergleich zu Placebo, ergänzt durch *Best-Supportive-Care* (BSC) bei Erwachsenen für die Behandlung von Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist, untersucht. In die Studie eingeschlossene Patient:innen wurden anhand der folgenden Patientencharakteristika, die zu Studienbeginn erhoben wurden, charakterisiert:

- **Demographische Charakteristika**
  - Alter
  - Geschlecht
  - Abstammung

- Body Mass Index (BMI)
- Geographische Region
- **Erkrankungscharakteristika**
  - Beta-Thalassämie-Diagnose (Beta-Thalassämie, HbE/Beta-Thalassämie, Beta-Thalassämie kombiniert mit Alpha-Thalassämie)
  - Beta-Thalassämie-Genmutation ( $\beta^0/\beta^0$ ,  $\beta^+/\beta^+$ ,  $\beta^+/\beta^0$  ohne Alpha-Thalassämie,  $\beta^0/\beta^0$ ,  $\beta^+/\beta^+$ ,  $\beta^+/\beta^0$  mit Alpha-Thalassämie,  $\beta^0/\beta$ ,  $\beta^+/\beta$  mit Alpha-Gen-Duplikation)
  - Hämoglobin-(Hb-)Wert zu Baseline ( $\geq 8,5$  g/dl,  $< 8,5$  g/dl; Stratifizierungsfaktor)
  - *Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes*-(NTDT-PRO-)Gesamtscore der Domäne Müdigkeit/Schwäche zu Baseline ( $\geq 3$  Punkte,  $< 3$  Punkte; Stratifizierungsfaktor)
  - Transfusionslast im Zeitraum vor der ersten Dosis (mind. 24 Wochen)
  - Vorhergehende Splenektomie
  - Lebereisenkonzentration zu Baseline (gemessen mittels Magnetresonanztomographie [MRT])
  - Serum-Ferritin-(sFerritin-)Wert zu Baseline
  - Eisenchelatherapie in den 24 Wochen vor Studienbeginn
  - *Eastern Cooperative Oncology Group* (ECOG) *Performance*-Status zu Baseline
  - 6-Minuten-Gehtest-(6MWT-)Messung zu Baseline
  - Fatigue-spezifische Skala (FACIT-Fatigue) des *Functional Assessment of Chronic Illness Therapy–Fatigue* (FACIT-F) zu Baseline

#### 4.2.5.2.2 Patientenrelevante Endpunkte

Als patientenrelevant wird verstanden, „wie eine Patientin oder ein Patient fühlt, seine Funktionen und Aktivitäten wahrnehmen kann oder ob sie oder er überlebt“ [2].

Die in der Studie BEYOND erhobenen patientenrelevanten Endpunkte sind im Nachfolgenden beschrieben. Sie gliedern sich entsprechend den Endpunktkategorien Mortalität, Morbidität,

Lebensqualität und Sicherheit und werden zur Ableitung des Zusatznutzens und zur Bewertung des Ausmaßes des Zusatznutzens herangezogen:

- **Mortalität**
  - Todesfälle
- **Morbidität**
  - Vermeidung von Transfusionen
  - Symptomatik gemäß NTDT-PRO
  - Symptomatik gemäß *Patient Global Impression of Severity* (PGI-S) und *Patient Global Impression of Change* (PGI-C)
  - Körperliche Leistungsfähigkeit gemäß 6MWT
- **Lebensqualität**
  - Lebensqualität gemäß *Short Form (36) Health Survey* (SF-36)
  - Lebensqualität gemäß FACIT-F
- **Sicherheit**
  - Unerwünschte Ereignisse (UE), insbesondere schwerwiegende UE (SUE), schwere UE (CTCAE-Grad  $\geq 3$ ) und zum Therapieabbruch führende UE, einschließlich einer Differenzierung nach SOC und PT gemäß MedDRA

Zusätzlich werden im Dossier u. a. Ergebnisse zur Regulation der Erythropoese, zur Kontrolle der Eisenüberladung und zu Hospitalisierungen dargestellt.

Es folgt eine detaillierte Beschreibung der einzelnen in der Studie BEYOND erhobenen Endpunkte mit Angaben zur Patientenrelevanz, Validität und Operationalisierung.

## **Mortalität**

### *Patientenrelevanz*

NTDT ist eine chronische Erkrankung, die nicht akut tödlich verläuft. Mortalität ist dennoch allgemein als patientenrelevanter Endpunkt anerkannt. Die Patientenrelevanz wurde vom G-BA im Beratungsgespräch zu Luspatercept bestätigt [4].

### *Validität und Operationalisierung*

Das Auftreten von Todesfällen stellt einen harten, etablierten Endpunkt dar, der objektiv erfasst werden kann und als valide einzustufen ist. Als Todesfälle wurden – unabhängig von der Todes-

ursache – alle während der Studie verstorbenen Patient:innen gezählt. Die Darstellung des Endpunkts erfolgt anhand des Anteils der Todesfälle pro Studienarm. Als Behandlungseffekt wird das unter Berücksichtigung der für die Randomisierung verwendeten Stratifizierungsfaktoren mittels Mantel-Haenszel-Methode geschätzte relative Risiko (RR) betrachtet.

## **Morbidität**

### ***Vermeidung von Transfusionen***

#### *Patientenrelevanz*

Das Kernmerkmal der Beta-Thalassämie ist ein Erythrozyten-Reifungs-Defekt und infolgedessen eine ineffektive Erythropoese, die eine Anämie mit Folgekomplikationen auslöst – darunter psychosoziale Probleme, Fatigue, extramedulläre Blutbildung und Organomegalie, gestörter Eisenstoffwechsel, Knochenverformungen, Wachstums- und Entwicklungsstörungen, bis hin zum frühen Tod. Die Gefahr einer extramedullären Blutbildung und die einhergehende Organomegalie ist bei Patient:innen mit NTDT besonders ausgeprägt, da eine ständige Kontrolle bei diesem Kollektiv in der Praxis schwierig sein kann [5, 6] (vgl. auch Modul 3 C). Weiterhin stellt der derzeitige Behandlungsstandard – eine bedarfsgerechte Transfusions-therapie mit Erythrozytenkonzentrat (EK) in Kombination mit Eisenchelatoren – nur eine unzureichende therapeutische Maßnahme dar. So führen Transfusionen nur kurzfristig zu einer Erhöhung des Hb-Spiegels, der einhergehend mit Anämie-Symptomen relativ schnell wieder abfallen kann. Somit kommt es durch Transfusionen zu keiner langfristigen Stabilisierung des Hb-Werts, die das Risiko von Folgekomplikationen durch längere anämische Perioden reduzieren würde. Auch NTDT-Patient:innen, die nur unregelmäßig oder selten EK-Transfusionen benötigen, können langfristig von einer Vermeidung dieser Transfusionen unter gleichzeitiger Linderung der Anämie profitieren, u. a. aufgrund der Vermeidung von Transfusionsrisiken (z. B. allergische Reaktionen, Alloimmunisierung, etc.), Folgekomplikationen, sowie einer weiteren Eisenüberladung, die sich in Form einer sekundären Hämosiderose wiederum auf Morbidität und Mortalität der Patient:innen auswirkt. Je länger ein stabiler Hb-Wert und eine verringerte Anämie erhalten werden kann, ohne auf Transfusionen angewiesen zu sein, desto länger kann der mögliche Übergang in eine transfusionsabhängige Beta-Thalassämie, die mit regelmäßiger Transfusionstherapie und einem erhöhten Behandlungsaufwand, insbesondere einem zeitlichen Aufwand einhergeht, verhindert werden. Auch unter Berücksichtigung des chronischen Hintergrunds der NTDT hat das Erreichen einer langfristigen Vermeidung von Transfusionen unter gleichzeitiger Steigerung und langfristigen Stabilisierung des Hb-Werts und somit einer anhaltenden Linderung der Anämie insgesamt eine wichtige Bedeutung für die Betroffenen. Zudem weist die Bundesärztekammer (BÄK) in ihrer Querschnitts-Leitlinie zur Therapie mit Blutkomponenten und Plasmaderivaten auf die Relevanz eines restriktiven Einsatzes von EK-Transfusionen hin. Wenn möglich, sind kausale oder andere, gleichwertige Therapien einer Transfusionstherapie vorzuziehen [24]. Somit führt eine langfristige Transfusionsfreiheit, die mit einer Stabilisierung oder sogar Verbesserung der Hb-Werte einhergeht, zu einer für die Patient:innen spürbaren Erleichterung und einer Verbesserung des therapielevanten Nutzens.

### *Validität und Operationalisierung*

Die Vermeidung von Transfusionen stellt einen harten, etablierten Endpunkt dar, der objektiv erfasst werden kann und als valide einzustufen ist. Die Vermeidung von Transfusionen wird im vorliegenden Dossier in erster Linie anhand der Zeit bis zur ersten Transfusion betrachtet. Ergänzend wird u. a. der Anteil der Patient:innen betrachtet, die eine Transfusionsfreiheit über mindestens 24 und mindestens 48 Wochen erreichten.

Die Zeit bis zur ersten Transfusion ist definiert als die Zeit von der ersten Dosis der Studienmedikation bis zur ersten EK-Transfusion während der Behandlungsphase. Die Auswertung erfolgte anhand einer Ereigniszeitanalyse. Zur Quantifizierung des Behandlungseffekts wurden das Hazard Ratio (HR) mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne eine EK-Transfusion während der Behandlungsphase wurden zum Zeitpunkt des betrachteten Datenschnitts bzw. 20 Tage nach der letzten Dosis oder bei Studienabbruch/Tod, je nachdem welches Ereignis früher eintrat, zensiert.

Weitere ergänzende Analysen (einschließlich Auswertungen zur Regulation der Erythropoese, insbesondere zur Veränderung der Hb-Werte) sind in Abschnitt 4.2.5.4 beschrieben.

### ***Kontrolle der Eisenüberladung (ergänzend dargestellt)***

Die Kontrolle der Eisenüberladung stellt in der vorliegenden Indikation einen klinisch relevanten Endpunkt dar und wird somit der Vollständigkeit halber im Dossier dargestellt, jedoch nicht zur Ableitung des Zusatznutzens herangezogen. Die Kontrolle der Eisenüberladung wird in erster Linie anhand der Änderung des sFerritin-Spiegels bewertet.

### *Klinische Relevanz*

Die durch NTDT bedingte erhöhte Eisenaufnahme führt unbehandelt zu Eisenablagerungen in verschiedenen Organen wie der Leber und den endokrinen Organen. Dadurch kommt es im Laufe der Zeit zu schweren Organschäden bis hin zum Tod (Siderose-bedingte Morbidität und Mortalität) [7-12]. Um einer Eisenüberladung adäquat entgegen zu wirken, sollten NTDT-Patient:innen engmaschig kontrolliert und ggf. mit einer Eisenchelatherapie behandelt werden [10, 11]. Auch wurde in der Studie BEYOND eine bedarfsgerechte Eisenchelatherapie im Rahmen der BSC in beiden Studienarmen sichergestellt. Entsprechend waren Unterschiede zwischen den Studienarmen hinsichtlich der Symptome einer Eisenüberladung nicht zu erwarten. Stattdessen wird eine Bewertung der Eisenüberladung anhand des klinisch relevanten Laborparameters sFerritin vorgenommen, der auch in der klinischen Praxis zur Verlaufskontrolle des Eisenhaushalts bei Patient:innen mit NTDT und zur Identifizierung von Trends hinsichtlich der Eisenlast im Körper empfohlen wird [10, 11].

### *Operationalisierung*

Im Dossier wird die mittlere Änderung des sFerritin-Spiegels in den letzten 24 Wochen der Behandlungsphase bis Woche 48 gegenüber den 24 Wochen vor der ersten Dosis (= Baseline-Wert) betrachtet. Die Auswertung der Änderung des sFerritin-Spiegels erfolgte mittels *Analysis of Covariance* (ANCOVA), wobei eine Adjustierung für den Baseline-Wert sowie den für die Randomisierung verwendeten Stratifizierungsfaktoren vorgenommen wurde. Zur Quantifi-

zierung des Gruppenunterschieds wird die adjustierte Mittelwertdifferenz (MWD) sowie der entsprechende Schätzer von Hedges'  $g$  dargestellt.

Als weitere Analysen werden die Änderung des Parameters LIC (*liver iron concentration*) sowie des Eisenchelator-Bedarfs dargestellt. Details hierzu finden sich in Abschnitt 4.2.5.4.

### ***Symptomatik gemäß NTDT-PRO (krankheitsspezifischer Fragebogen)***

#### *Patientenrelevanz*

Das Messinstrument NTDT-PRO ist ein patientenberichteter Fragebogen, der die Wahrnehmung der Patient:innen bezüglich ihrer Thalassämie-bedingten Symptomatik misst. Es handelt sich somit um einen patientenrelevanten Endpunkt, da er die durch die Patient:innen direkt wahrgenommene Belastung abbildet.

#### *Validität und Operationalisierung*

Der NTDT-PRO wurde für Patient:innen mit NTDT entwickelt und stellt den derzeit einzigen krankheitsspezifischen Fragenbogen zur Erhebung Anämie-bezogener Symptome in diesem Patientenkollektiv dar. Der Fragebogen wurde unter anderem mit Hilfe von *concept elicitation* und kognitiven Interviews sowie der Unterstützung von klinischen Experten im Rahmen des Zulassungsprozesses von Luspatercept für NTDT entwickelt [13].

Der NTDT-PRO besteht aus einem elektronischen Tagebuch, in dem der Schweregrad der Beta-Thalassämie-bezogenen Symptome während der letzten 24 Stunden dokumentiert wird. Der Fragebogen umfasst insgesamt sechs Items, mit denen der Schweregrad bestimmter Symptome auf einer numerischen Rating-Skala von 0 – 10 („keine“ bis „extrem/sehr stark“) bewertet wird:

- Müdigkeit (Energienmangel) ohne körperliche Belastung
- Müdigkeit (Energienmangel) während körperlicher Belastung
- Schwäche (Kraftlosigkeit) ohne körperliche Belastung
- Schwäche (Kraftlosigkeit) während körperlicher Belastung
- Kurzatmigkeit ohne körperliche Belastung
- Kurzatmigkeit während körperlicher Belastung

Aus den Einzelitems werden die Gesamtscores für zwei verschiedene Domänen bestimmt, die ebenfalls eine Skalenspannweite von 0 bis 10 aufweisen:

- Gesamtscore in der Domäne Müdigkeit/Schwäche (*tiredness/weakness*; T/W)
- Gesamtscore in der Domäne Kurzatmigkeit (*shortness of breath*, SoB)

Der Gesamtscore in der Domäne Müdigkeit/Schwäche wird als der Mittelwert der vier Items zu Müdigkeit und Schwäche berechnet; der Gesamtscore in der Domäne Kurzatmigkeit wird als Mittelwert der beiden Items zu Kurzatmigkeit berechnet.

Der NTDT-PRO wurde in einer prospektiven Beobachtungsstudie validiert und zeigt gute psychometrische Eigenschaften [14]. Auch der G-BA bestätigt die überwiegend gute Reliabilität und Validität des NTDT-PRO [4]. Um die Konstruktvalidität des NTDT-PRO zu untersuchen, wurden Korrelationen zwischen den zwei NTDT-PRO Gesamtscores und den Subskalen etablierter Instrumente (SF-36, Functional Assessment of Cancer Therapy – Anemia [FACT-An]) zu verschiedenen Zeitpunkten der Studie (Woche 1, 3 und 12) mittels Spearman Korrelationskoeffizient ausgewertet. Im Ergebnis bestehen teils hohe Korrelationen zu vergleichbaren Domänen der etablierten Instrumente (wobei es sich durchgängig um negative Korrelationen handelt, da hohe Werte in den NTDT-PRO-Skalen im Gegensatz zum SF-36 und FACT-An eine hohe Belastung anzeigen). So wurden für den NTDT-PRO-T/W-Gesamtscore mit der SF-36-Subskala zur Vitalität für die verschiedenen Zeitpunkte Spearman Korrelationskoeffizienten von  $r = -0,45$  bis  $-0,53$  berichtet; mit der Anämie-spezifischen Skala des FACT-An und der ebenfalls im Rahmen des FACT-An erhobenen Fatigue-spezifischen Skala (FACIT-Fatigue Skala) lagen die Korrelationskoeffizienten in der Spanne von  $r = -0,34$  bis  $-0,80$ . Für den NTDT-PRO-SoB-Gesamtscore wurden mit den SF-36 Skalen zur körperlichen Funktionsfähigkeit, zur Vitalität und zur allgemeinen Gesundheitswahrnehmung Spearman Korrelationskoeffizienten in der Spanne von  $r = -0,37$  bis  $-0,46$  angegeben; mit der FACIT-Fatigue-Skala lagen die Korrelationskoeffizienten zwischen  $r = -0,38$  und  $-0,58$ . Darüber hinaus wurden für die Gesamtscores des NTDT-PRO in beiden Domänen zu allen Zeitpunkten jeweils eine gute Test-Retest-Reliabilität (*Intraclass Correlation Coefficient* [ICC] = 0,88 bis 0,98), sowie eine gute interne Konsistenz mit einem Cronbachschen Alpha von 0,91 für den NTDT-PRO-T/W-Gesamtscore und einem Spearman-Brown-Koeffizienten von 0,78 für den NTDT-PRO-SoB-Gesamtscore festgestellt.

Die Auswertung erfolgte für beide Gesamtscores und die Einzelitems jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber dem Ausgangswert. Als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung werden dabei 1,5 Punkte betrachtet (entsprechend einer Veränderung von 15 % der Skalenspannweite).

Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell mit Adjustierung für den Baseline-Wert sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert.

Weitere Analysen (u. a. eine Auswertung mittels MMRM) sind in Abschnitt 4.2.5.4 beschrieben

### ***Symptomatik gemäß PGI-S und PGI-C (patientenberichtete Symptomschwere)***

#### *Patientenrelevanz*

Die Messinstrumente *Patient Global Impression of Severity* (PGI-S) und *Patient Global Impression of Change* (PGI-C) sind patientenberichtete Fragebögen, die die Wahrnehmung der Patient:innen bezüglich ihrer Thalassämie-bedingten Symptomatik messen. Es handelt sich somit um patientenrelevante Endpunkte, da sie für die Patient:innen direkt wahrnehmbare Veränderungen abbilden. Dies hat der G-BA im Beratungsgespräch bestätigt [4].

#### *Validität und Operationalisierung*

Der PGI-S besteht aus einer einzigen Frage, die die Patient:innen bittet, die Schwere ihrer Thalassämie-bedingten Symptomatik anhand einer 11 Punkte umfassenden numerischen Rating-Skala (von 0 = „keine Symptome“ bis 10 = „sehr schwere Symptome“) zu bewerten. Auch der PGI-C besteht aus einer einzigen Frage, wobei die Patient:innen hier angehalten werden auf einer 7 Punkte umfassenden verbalen Rating-Skala die gesamte Änderung ihrer Thalassämie-bedingten Symptomatik seit Beginn der Studie einzuschätzen („sehr viel schlechter“, „viel schlechter“, „ein wenig schlechter“, „unverändert“, „ein wenig besser“, „viel besser“, „sehr viel besser“).

Die Validität und die Operationalisierung des PGI-S und PGI-C wurden bereits in vorherigen Verfahren vom G-BA anerkannt und für die Nutzenbewertung berücksichtigt [15-17]. Die Symptomatik gemäß PGI-S und PGI-C wird somit als valide und zur Ableitung eines Zusatznutzens geeignet eingestuft.

Die Auswertung erfolgte wie bei den anderen patientenberichteten Endpunkten (vgl. z. B. NTDT-PRO) jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber Baseline. Beim PGI-S wurden 1,5 Punkte (entsprechend 15 % der Skalenspannweite) als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung herangezogen. Beim PGI-C wurden die beiden besten Bewertungen der Veränderung („viel besser“ und „sehr viel besser“) als bewertungsrelevante Verbesserung und die beiden schlechtesten Bewertungen („viel schlechter“ und „sehr viel schlechter“) als bewertungsrelevante Verschlechterung betrachtet.

Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Beim PGI-S wurde der Baseline-Score als Kovariate im Cox-Modell berücksichtigt. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert.

Weitere Analysen (u. a. eine Auswertung des PGI-S mittels MMRM) sind in Abschnitt 4.2.5.4 beschrieben.

### ***Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)***

#### *Patientenrelevanz*

Der 6MWT ist ein funktioneller Test zur Bestimmung der Gehstrecke, die Patient:innen innerhalb von sechs Minuten auf ebenem Terrain zurücklegen können. Während des Tests dürfen die Patient:innen ihr Lauftempo anpassen oder Pausen einlegen, werden jedoch gebeten die ihnen innerhalb der sechs Minuten maximal mögliche Strecke zurückzulegen. Die Anämie kann bei Patient:innen mit einer NTDT zu erheblichen Einschränkungen der körperlichen Funktion, Belastbarkeit und Ausdauer führen. Der 6MWT macht diese Einschränkungen messbar und gibt somit Aufschluss darüber, wie stark Patient:innen bei entsprechenden Aktivitäten des täglichen Lebens beeinträchtigt sind. Der G-BA hat den 6MWT als patientenrelevantes Instrument zur Messung der körperlichen Belastbarkeit bei NTDT und anderen Erkrankungen eingestuft [4, 18].

#### *Validität und Operationalisierung:*

Der 6MWT ist ein standardisierter und etablierter Test zur Bestimmung der körperlichen Belastbarkeit. Er wird als einfach durchzuführender, reproduzierbarer Test zur Diagnostik und Verlaufskontrolle unterschiedlicher Indikationen (z. B. schwere Herzinsuffizienz, pulmonale arterielle Hypertonie) eingesetzt. Somit ist der 6MWT auch zur Messung der körperlichen Belastbarkeit bei NTDT geeignet.

Die Auswertung des 6MWT erfolgte mittels ANCOVA basierend auf der Änderung zu Woche 48 gegenüber Baseline, wobei eine Adjustierung für den Baseline-Wert sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren vorgenommen wurde. Zur Quantifizierung des Gruppenunterschieds wird die adjustierte Mittelwertdifferenz sowie der entsprechende Schätzer von Hedges'  $g$  betrachtet.

Ergänzend wird die Änderung des 6MWT zu Woche 24 dargestellt.

### **Lebensqualität**

#### *Patientenrelevanz*

Endpunkte, die die gesundheitsbezogene Lebensqualität der Patient:innen abbilden, sind grundsätzlich patientenrelevant. Aufgrund der chronischen Natur ihrer Erkrankung ist die gesundheitsbezogene Lebensqualität bei Patient:innen mit NTDT darüber hinaus von besonderer Relevanz.

Zur Erhebung der Lebensqualität empfiehlt der G-BA die Verwendung sowohl eines generischen als auch eines krankheitsspezifischen Fragebogens [4].

#### ***Lebensqualität gemäß SF-36 (generischer Fragebogen)***

##### *Validität und Operationalisierung*

Der SF-36 ist ein validierter und etablierter Fragebogen zur Erhebung der gesundheitsbezogenen Lebensqualität und wurde vom G-BA bereits als valider, patientenrelevanter

Endpunkt anerkannt [19-22]. Auch bei Patient:innen mit Beta-Thalassämie wurde dieser Fragebogen bereits in verschiedenen Studien herangezogen, um die Lebensqualität der Patient:innen abzubilden [23-26]. Im Beratungsgespräch zu Luspatercept wurde der SF-36 vom G-BA als generischer Fragebogen empfohlen [4].

In der Studie BEYOND wurde die zweite Version des SF-36 verwendet (SF36v2). Der Fragebogen umfasst 36 Items, aus denen sich acht verschiedene Domänen bestimmen lassen (körperliche Funktionsfähigkeit, körperliche Rollenfunktion, körperliche Schmerzen, allgemeine Gesundheitswahrnehmung, Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion und psychisches Wohlbefinden). Diese wiederum werden unter Verwendung spezieller Gewichte zu zwei Summenscores zum körperlichen Wohlbefinden (*Physical Component Summary*; PCS) und zum psychischen Wohlbefinden (*Mental Component Summary*; MCS) zusammengefasst. Diese Scores sind derart transformiert, dass sich für die Normalbevölkerung (ermittelt durch eine Normstichprobe) jeweils ein Mittelwert von 50 und eine Standardabweichung von 10 ergäbe. Somit bedeuten Werte über 50 eine überdurchschnittliche und Werte unter 50 eine unterdurchschnittliche Lebensqualität.

Die Auswertung erfolgte wie bei den anderen patientenberichteten Endpunkten (vgl. z. B. NTDT-PRO) jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber Baseline. Als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung werden dabei für den PCS 9,4 Punkte und für den MCS 9,6 Punkte betrachtet. Den beobachteten Minimal- und Maximalwerten der Normstichprobe von 2009 entsprechend bilden diese Werte jeweils die vom G-BA geforderte Responseschwelle von 15 % der Skalenspannweite ab [27].

Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell mit Adjustierung für den Baseline-Wert sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert.

Weitere Analysen (u. a. eine Auswertung mittels MMRM) sind in Abschnitt 4.2.5.4 beschrieben.

### ***Lebensqualität gemäß FACIT-F (Fatigue-spezifischer Fragebogen)***

#### *Validität und Operationalisierung*

Zu den Symptomen der Anämie, bedingt durch die Beta-Thalassämie, zählen vor allem Belastungsintoleranzen, die zu einer Einschränkung der Leistungsfähigkeit und der Lebensqualität führen können. Diese Einschränkungen können mit dem symptom-spezifischen Instrument FACIT-F erhoben werden, das die gesundheitsbezogene Lebensqualität sowie den Einfluss der Fatigue auf die täglichen Aktivitäten der Patient:innen erfasst. Der FACIT-F ist ein validiertes Selbsterhebungsinstrument, das für Patient:innen mit chronischen Erkrankungen

entwickelt wurde. Somit ist eine Übertragbarkeit der Validität und Interpretierbarkeit für Patient:innen mit NTDT gegeben. Auch in früheren Nutzenbewertungsverfahren in verschiedenen Indikationen wurde der FACIT-F vom G-BA bereits als valide und patientenrelevant anerkannt [28-31]. Bei Fatigue handelt es sich um eines der häufigsten Symptome bei NTDT-Patient:innen mit Anämie [32, 33]. Die Verwendung des Fatigue-spezifischen Fragebogens FACIT-F für diese Patient:innen ist daher sachgerecht.

Der FACIT-F Fragebogen umfasst insgesamt 40 Items, die durch die Patient:innen jeweils auf einer Fünf-Punkte-Skala von 0 („überhaupt nicht“) bis 4 („sehr“) bewertet werden. Der Fragebogen ist modular aufgebaut und gliedert sich in einen 27 Fragen umfassenden generischen Kern (FACT-G), bestehend aus den Subskalen physisches Wohlbefinden (PWB), soziales/familiäres Wohlbefinden (SWB), emotionales Wohlbefinden (EWB) und funktionelles Wohlbefinden (FWB) und einer 13 Fragen umfassenden Fatigue-spezifischen Skala (FACIT-Fatigue Skala), die das individuelle Niveau der Fatigue während der üblichen Tagesaktivitäten in der vergangenen Woche erfasst. Die Gesamtscores des FACT-G und FACIT-F werden jeweils als Summenscores aus den zugehörigen Subskalen berechnet. Ein höherer Wert entspricht sowohl bei den Subskalen als auch bei den Summenscores einer besseren Lebensqualität.

Die Auswertung erfolgte wie bei den anderen patientenberichteten Endpunkten (vgl. z. B. NTDT-PRO) jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber dem Ausgangswert. Als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung werden dabei 15 % der jeweiligen Skalenweite betrachtet (siehe Tabelle 4-3).

Tabelle 4-3: Übersicht der herangezogenen Responseschwellen für den Endpunkt gesundheitsbezogene Lebensqualität gemessen anhand des FACIT-F

Domäne/Summenscore	Skalenweite	Responseschwelle (15 % der Skalenweite)
FACIT-F Gesamtscore	0 – 160	24
FACT-G Gesamtscore	0 – 108	17
Physisches Wohlbefinden (PWB)	0 – 28	5
Soziales/familiäres Wohlbefinden (SWB)	0 – 28	5
Emotionales Wohlbefinden (EWB)	0 – 24	4
Funktionelles Wohlbefinden (FWB)	0 – 28	5
Fatigue-spezifische Skala (FACIT-Fatigue Skala)	0 – 52	8

Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell mit Adjustierung für den Baseline-Wert sowie ein stratifizierter Log-Rank-

Test berechnet. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert.

Weitere Analysen (u. a. eine Auswertung mittels MMRM) sind in Abschnitt 4.2.5.4 beschrieben.

## **Sicherheit**

### ***Unerwünschte Ereignisse***

#### *Patientenrelevanz*

Die Erhebung von UE, die im zeitlichen Zusammenhang mit der Therapie auftreten, ist ein wesentlicher Bestandteil der Bewertung der Sicherheit und Verträglichkeit eines Arzneimittels. Im Allgemeinen ist das Auftreten von UE, insbesondere von schweren und schwerwiegenden UE, für die Patient:innen unmittelbar erfahrbar und gilt als patientenrelevant.

#### *Validität und Operationalisierung*

Das Auftreten von UE stellt einen harten, etablierten Endpunkt dar, der zumindest für schwere und schwerwiegende UE weitgehend objektiv erfasst werden kann und als valide einzustufen ist. Die Operationalisierung der UE in der Studie BEYOND beinhaltet alle UE, die ab dem Tag der ersten Dosis der Studienmedikation bis 63 Tage nach der letzten Dosis auftraten. Die Erfassung der UE folgte den internationalen Standards. Als UE wurden, unabhängig von einem kausalen Zusammenhang mit der Studienmedikation, alle schädlichen, unbeabsichtigten oder unerwünschten medizinischen Ereignisse betrachtet, die bei Studienbeginn noch nicht vorlagen oder sich während der Studie verschlimmerten.

Die Kodierung der UE wurde mittels des *Medical Dictionary for Regulatory Activities* (MedDRA) Version 23.0 durchgeführt. Der Schweregrad der UE wurde nach den *Common Terminology Criteria for Adverse Events* (CTCAE) in der Version 4.0 des National Cancer Institute (NCI) klassifiziert [34]. UE, deren Schweregrad durch die CTCAE-Klassifizierung nicht definiert war, wurden durch den oder die Prüfarzt:in anhand einer 5-Punkte-Skala bewertet (Grad 1: mild; Grad 2: moderat; Grad 3: schwer; Grad 4: lebensbedrohlich; Grad 5: tödlich).

Als schwerwiegende UE (SUE) wurden solche Ereignisse berücksichtigt, die

- tödlich verliefen,
- lebensbedrohlich waren,
- eine Hospitalisierung erforderlich machten oder zur Verlängerung eines bestehenden stationären Aufenthalts führten,
- in bestehenden oder bedeutsamen Behinderungen/Unfähigkeiten resultierten,

- angeborene Anomalien/Geburtsdefekte nach sich zogen, oder
- wichtige medizinische Ereignisse waren, die nicht unmittelbar tödlich oder lebensbedrohlich waren oder eine Hospitalisierung oder Behinderung verursachten, aber den Patienten bzw. die Patientin diesen Risiken aussetzen oder eine medizinische oder chirurgische Behandlung zur Prävention dieser Ereignisse nötig machten.

Im Rahmen der Nutzenbewertung ist gemäß Vorgaben des G-BA eine Analyse der UE ohne Berücksichtigung von Ereignissen, die auf die Grunderkrankung zurückzuführen sind, vorzulegen. Um entsprechende Ereignisse zu identifizieren, wurde auf Ebene der *Preferred Terms* (PT) ein medizinischer Review aller in der Studie BEYOND aufgetretenen Ereignisse vorgenommen. Hierbei wurden jedoch keine Ereignisse identifiziert, bei denen eine alternative Ätiologie (z. B. eine sich verschlimmernde Vorerkrankung oder die Studientherapie) mit hinreichender Sicherheit ausgeschlossen werden konnte. Somit wurden bei den Auswertungen im Dossier alle UE ohne post-hoc Löschung bestimmter Terme berücksichtigt.

Im Rahmen des Dossiers werden folgende UE-Kategorien betrachtet („UE-Hauptkategorien“):

- Jegliche UE (ergänzend dargestellt)
- SUE
- Schwere UE (CTCAE-Grad  $\geq 3$ )
- Zum Therapieabbruch führende UE

Darüber hinaus werden die UE dieser Kategorien differenziert auf Ebene der *System Organ Class* (SOC) und PT dargestellt.

Die Auswertung erfolgte jeweils anhand einer Ereigniszeitanalyse auf Basis der Zeit von der ersten Dosis bis zum ersten UE innerhalb der betrachteten Kategorie. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne ein entsprechendes UE wurden 63 Tage nach der letzten Dosis bzw. zum Zeitpunkt des betrachteten Datenschnitts oder bei vorherigem Studienabbruch/Tod, je nachdem, welches Ereignis früher eintrat, zensiert.

Weitere spezifische UE (u. a. UE von besonderem Interesse) und ergänzende Analysen (in Bezug auf Hospitalisierungen) werden im Rahmen von Zusatzanalysen dargestellt (siehe Abschnitt 4.2.5.4).

#### 4.2.5.2.3 Verwendete statistische Methodik

##### *Ereigniszeitanalysen*

Ereigniszeitanalysen wurden stets anhand eines stratifizierten Cox-Regressionsmodell und eines stratifizierten Log-Rank-Tests durchgeführt. Zur Quantifizierung des Behandlungseffekts wird das HR aus dem Cox-Regressionsmodell herangezogen. Die Stratifizierung erfolgte

basierend auf den Stratifizierungsfaktoren der Randomisierung (Baseline-Hb-Wert  $\geq 8,5$  g/dl vs.  $< 8,5$  g/dl] und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche  $\geq 3$  Punkte vs.  $< 3$  Punkte]). Für die Analyse der Zeit bis zur Verbesserung oder Verschlechterung bei den patientenberichteten Endpunkten wurde neben der Behandlung auch der Baseline-Wert des betrachteten Scores als unabhängige Variable im Cox-Regressionsmodell berücksichtigt. Zusätzlich werden Ereigniszeitkurven gemäß Kaplan-Meier-Methode dargestellt und die entsprechende mediane Zeit bis zum Ereignis pro Studienarm berichtet.

### ***Analysen mit binären Messwerten***

Für dichotome Zielvariablen wird der absolute und prozentuale Anteil der Patient:innen mit einem Ereignis (z. B. Ansprechen) pro Studienarm berichtet. Als Behandlungseffekt wird der Mantel-Haenszel-Schätzer des relativen Risikos (RR) mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung (Baseline-Hb-Wert  $\geq 8,5$  g/dl vs.  $< 8,5$  g/dl] und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche  $\geq 3$  Punkte vs.  $< 3$  Punkte]) verwendet. Der zugehörige p-Wert wurde mittels Normal-Approximation bestimmt. Zusätzlich werden die Mantel-Haenszel-Schätzer des Odds Ratios (OR) und der Risikodifferenz (RD) dargestellt.

### ***Analysen mit stetigen Messwerten***

#### ***ANCOVA***

Für stetige Variablen wurde die Veränderung gegenüber Baseline mittels ANCOVA (*analysis of covariance*) ausgewertet. Im entsprechenden Modell wurden, neben dem Baseline-Wert und dem Behandlungsarm, die für die Randomisierung verwendeten Stratifizierungsfaktoren (Baseline-Hb-Wert  $\geq 8,5$  g/dl vs.  $< 8,5$  g/dl] und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche  $\geq 3$  Punkte vs.  $< 3$  Punkte]) als unabhängige Variablen berücksichtigt. Zur Quantifizierung des Behandlungseffekts wird der adjustierte Modellschätzer der Mittelwertdifferenz sowie die hieraus abgeleitete standardisierte Mittelwertdifferenz (SMD) in Form von Hedges' g dargestellt.

Um im Zweifelsfall die Relevanz bei Vorliegen eines signifikanten Effekts zu beurteilen, wird je nach Effektrichtung eine Irrelevanzschwelle von 0,2 bzw. -0,2 für das 95 %-KI von Hedges' g (SMD) verwendet. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als „klein“ angesehen werden kann [2].

#### ***MMRM***

Als Zusatzanalyse für die patientenberichteten Endpunkte wurde zudem die Veränderung gegenüber Baseline mittels MMRM ausgewertet. Details zu den MMRM-Analysen finden sich bei der Beschreibung der Zusatzanalysen in Abschnitt 4.2.5.4.

### ***Analysepopulationen***

Im vorliegenden Dossier werden die folgenden beiden Analysepopulationen betrachtet:

- ITT-Population: Diese Population umfasst gemäß statistischem Analyseplan alle randomisierten Patient:innen der Gruppenzuteilung gemäß Randomisierung

entsprechend, unabhängig davon, welche Studienmedikation die Patient:innen tatsächlich erhalten haben.

- **Safety-Population:** Diese Population umfasst gemäß statistischem Analyseplan alle randomisierten Patient:innen, die mindestens eine Dosis Studienmedikation erhalten haben. Patient:innen werden dem Behandlungsarm zugeordnet, dessen Behandlung sie tatsächlich erhalten haben.

Die im vorliegenden Dossier präsentierten Auswertungen der Patientencharakteristika und der Wirksamkeitsendpunkte (einschließlich Fragebögen zu Symptomen und Lebensqualität) beruhen auf der ITT-Population. Für Sicherheitsendpunkte wurde, sofern nicht anders gekennzeichnet, dagegen die Safety-Population verwendet. Im vorliegenden Fall unterscheiden sich die beiden Populationen jedoch nicht (vgl. Abschnitt 4.3.1.2.1).

### Datenbasis

Die Auswertungen im vorliegenden Dossier beruhen auf dem primären Datenschnitt der Studie BEYOND vom 14.09.2020 (Datenbankschluss 23.11.2020). Hierbei handelt es sich um den präspezifizierten Datenschnitt für die primäre Analyse der Studie, die vor Entblindung der Studie durchgeführt wurde, nachdem alle Patient:innen die 48-wöchige primäre Behandlungsphase beendet hatten.

Daten aus der nachfolgenden offenen Behandlungsphase der Studie sind für die Beantwortung der Fragestellung der Nutzenbewertung nicht geeignet, da in dieser Phase entsprechend einer Empfehlung des *Data Monitoring Committee* (DMC) die Patient:innen aus beiden Behandlungsarmen eine Therapie mit Luspatercept erhalten konnten (Therapiewechsel im Placebo-Arm).

### 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 Heterogenitätsschätzung<sup>6</sup> erfolgen. Im

---

<sup>6</sup> 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.

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 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 verwendet<sup>7</sup> 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<sup>8, 6</sup> 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. Es liegt nur eine relevante Studie vor, sodass keine Meta-Analysen durchgeführt wurden.

#### **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

---

<sup>7</sup> 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.

<sup>8</sup> Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-560.

Effektmaßes. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ soll für Sensitivitätsanalysen verwendet werden.

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

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

Im Folgenden werden abseits klassischer Sensitivitätsanalysen zur Untersuchung der methodischen Robustheit auch sonstige ergänzende Auswertungen (z. B. alternative Operationalisierungen oder supportive Endpunkte), die im Ergebnisteil als Zusatzanalysen präsentiert werden, beschrieben.

### **Vermeidung von Transfusionen**

Zusätzlich zur Hauptanalyse (der Zeit bis zur ersten Transfusion) werden im Dossier Responderanalysen zur Transfusionsfreiheit über  $\geq 24$  bzw.  $\geq 48$  Wochen und eine deskriptive Analyse zur Anzahl benötigter Transfusionen pro Patient:in betrachtet. Als weiterer ergänzender Aspekt wird die Regulation der Erythropoese anhand der mittleren Änderung der Hb-Werte, des mittleren Verlaufs der Hb-Werte, des erythroiden Ansprechens sowie der Dauer des erythroiden Ansprechens dargestellt.

#### *Responderanalysen zur Transfusionsfreiheit*

Wenngleich aufgrund der ungleichen Beobachtungsdauern der Studienarme die Operationalisierung mittels Ereigniszeitanalyse als geeigneter für die Nutzenbewertung angesehen wird, werden im Sinne einer umfassenden Darstellung der Vermeidung von Transfusionen im Dossier auch Responderanalysen zur Transfusionsfreiheit präsentiert.

Eine Transfusionsfreiheit von  $\geq 24$  Wochen liegt bei einer Patientin oder einem Patienten vor, wenn diese bzw. dieser während der betrachteten Behandlungsphase in einem kontinuierlichen Intervall von  $\geq 24$  Wochen keine EK-Transfusionen erhalten hat. Die Auswertung basiert auf dem Anteil der Patient:innen mit einer Transfusionsfreiheit über  $\geq 24$  Wochen während der Behandlungsphase im Zeitraum (i) bis Woche 24, (ii) bis Woche 48 und (iii) bis zum primären Datenschnitt. Analog hierzu erfolgte darüber hinaus eine Auswertung der Transfusionsfreiheit  $\geq 48$  Wochen während der Behandlungsphase im Zeitraum (i) bis Woche 48 und (ii) bis zum primären Datenschnitt.

Als Schätzer des Behandlungseffekts wird jeweils das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.

#### *Anzahl benötigter Transfusionen (deskriptive Analyse)*

Neben der Zeit bis zur ersten Transfusion und dem Erreichen einer Transfusionsfreiheit spielen auch die Anzahl benötigter Transfusionen eine wichtige Rolle und werden im Dossier berichtet.

Die Anzahl benötigter Transfusionen wird als kategorielle Variable anhand des Patientenanteils pro Kategorie deskriptiv ausgewertet.

### *Mittlere Änderung der Hb-Werte*

Um die Regulation bzw. Verbesserung der Erythropoese und Anämie zu bewerten, werden die Hb-Werte der Patient:innen ausgewertet. Der Hb-Spiegel stellt einen klinisch relevanten und therapiesteuernden Parameter im Kontext der Therapie der NTDT dar. Somit unterstreicht ein Anstieg der Hb-Werte den Therapieerfolg und belegt, dass eine potenzielle Vermeidung von Transfusionen nicht auf Kosten einer Verschlechterung der Erythropoese bzw. einer Verschlimmerung der Anämie erreicht wurde. Die Anämie – das Kernmerkmal der Beta-Thalassämie – ist gekennzeichnet durch pathologisch niedrige Hb-Werte, die mit einer hohen Rate verschiedener Komplikationen verbunden sind (darunter primäre Hämosiderose, Organomegalie, Wachstums- und Entwicklungsstörungen, sowie Knochenveränderungen, insbesondere Verformungen und Osteopenie-/Osteoporose-Syndrome) und sich negativ auf die Lebensqualität der Patient:innen auswirken [5]. Eine Korrektur der gestörten Erythropoese durch medikamentöse Interventionen, wie Luspatercept, führt zu einem Anstieg des Hb-Spiegels, wodurch einer Anämie sowie deren assoziierten Symptomen (wie Belastungsdyspnoe, Tachykardie und Ermüdung/Fatigue) vorgebeugt werden kann [35]. Weiterhin wird mit einer langfristigen Stabilisierung des Hb-Spiegels nicht nur eine langfristige Transfusionsfreiheit sichergestellt, sondern es werden auch anämische Perioden mit einhergehenden Folgekomplikationen vermieden. So ist ein mittlerer Hb-Anstieg von  $\geq 1,0$  g/dl mit einer klinisch bedeutsamen Reduktion von Symptomen und Komplikationen verbunden [36]. Außerdem dient der Hb-Wert u. a. auch zur klinischen Bewertung der anhaltenden Verbesserung der Erythropoese und Anämie. Aus diesem Grund handelt es sich bei einem Anstieg des Hb-Werts um  $\geq 1,0$  g/dl gegenüber Baseline in Abwesenheit von EK-Transfusionen um einen klinisch relevanten Endpunkt.

Um die Regulation bzw. Verbesserung der Erythropoese abzubilden, wird im vorliegenden Dossier in erster Linie die mittlere Änderung der Hb-Werte während der Behandlungsphase im Zeitraum bis Woche 48 gegenüber Baseline anhand einer ANCOVA ausgewertet, wobei als post-Baseline-Wert der Mittelwert aller Messungen während der Behandlungsphase bis Woche 48 betrachtet wird. Analog hierzu wird auch die mittlere Änderung der Hb-Werte während der Behandlungsphase im Zeitraum bis Woche 24 betrachtet. Als Baseline-Hb-Wert wurde der Mittelwert der Hb-Messungen innerhalb der vier Wochen vor der ersten Dosis herangezogen. In dem verwendeten Modell wurden als unabhängige Variablen der Baseline-Hb-Wert (als stetige Variable), der Behandlungsarm und mit dem Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche ( $\geq 3$  Punkte vs.  $<3$  Punkte) einer der beiden Stratifizierungsfaktoren der Randomisierung verwendet. Aufgrund der Berücksichtigung des Baseline-Hb-Werts als stetige Variable im Modell wurde der Baseline-Hb für diese Analyse nicht als weiterer Stratifizierungsfaktor ( $\geq 8,5$  g/dl vs.  $< 8,5$  g/dl) einbezogen.

### *Graphische Darstellung des mittleren Verlaufs der Hb-Werte*

Zur graphischen Darstellung des Behandlungsverlaufs wurden die Hb-Werte der Patient:innen anhand des arithmetischen Mittelwerts pro Behandlungsgruppe jeweils in aufeinander-

folgenden 3-Wochen-Intervallen ab Woche 2 (Woche 2–4, Woche 5–7, Woche 8–10, usw.) zusammengefasst. Bei Patient:innen mit mehr als einer Messung innerhalb eines Intervalls wurde der Durchschnitt dieser Messungen herangezogen.

#### *Erythroides Ansprechen*

Als weitere Zusatzanalyse zur Verbesserung der Hb-Werte wird im vorliegenden Dossier der Anteil der Patient:innen betrachtet, die während der Behandlungsphase im Mittelwert über 12 Wochen einen Anstieg des Hb-Spiegels um  $\geq 1,0$  g/dl im Vergleich zum mittleren Hb-Wert in den 4 Wochen vor der ersten Dosis erreichten (im vorliegenden Dossier als „erythroides Ansprechen“ bezeichnet). Für längere Zeiträume wurde der Hb-Anstieg in Form eines gleitenden Mittelwerts über 12 Wochen (= 84 Tage) berechnet. D. h. ein erythroides Ansprechen über einen bestimmten Zeitraum (z. B. 24 Wochen) liegt vor, wenn der Mittelwert der Hb-Messungen einer Patientin oder eines Patienten in jedem möglichen 12-Wochen-Intervall innerhalb dieses Zeitraums (also von Tag x bis x+84, von Tag x+1 bis x+85, usw.) im Vergleich zum mittleren Hb-Wert in den 4 Wochen vor der ersten Dosis (= Baseline-Hb-Wert) um  $\geq 1,0$  g/dl erhöht war.

Die Auswertung basiert auf dem Anteil der Patient:innen mit einem erythroiden Ansprechen  $\geq 24$  Wochen während der Behandlungsphase im Zeitraum (i) bis Woche 24, (ii) bis Woche 48 und (iii) bis zum primären Datenschnitt. Analog hierzu erfolgte darüber hinaus eine Auswertung des erythroiden Ansprechens  $\geq 48$  Wochen während der Behandlungsphase im Zeitraum (i) bis Woche 48 und (ii) bis zum primären Datenschnitt.

Als Behandlungseffekt wird jeweils das unter Berücksichtigung der für die Randomisierung verwendeten Stratifizierungsfaktoren mittels Mantel-Haenszel-Methode geschätzte relative Risiko betrachtet.

#### *Längste Dauer des erythroiden Ansprechens*

Um Unterschiede in der Dauer des erythroiden Ansprechens differenziert abzubilden, wurde zusätzlich die längste Dauer des Ansprechens anhand einer Ereigniszeitanalyse untersucht. Hierbei wurden nur Patient:innen einbezogen, die während der Behandlungsphase mindestens einmal ein erythroides Ansprechen erreichten. Die längste Dauer des erythroiden Ansprechens entspricht dem längsten Zeitraum, in dem die Hb-Werte einer Patientin bzw. eines Patienten im gleitenden Mittelwert über 12 Wochen durchgängig um  $\geq 1,0$  g/dl erhöht waren. Ein Ereignis liegt bei Patient:innen vor, die ihr längstes erythroides Ansprechen bis zum Datenschnitt bzw. bis 20 Tage nach Therapieabbruch oder Studienabbruch/Tod (je nachdem, was früher eintrat) wieder verloren hatten. Patient:innen, deren erythroides Ansprechen zu diesem Zeitpunkt noch andauerte, wurden zensiert.

#### **Kontrolle der Eisenüberladung**

Ergänzend zur Änderung des sFerritin-Spiegels werden im Dossier die Änderung des Parameters LIC sowie die Änderung des Eisenchelator-Bedarfs dargestellt.

### *Änderung des Parameters LIC*

Die LIC (*liver iron concentration*) bildet direkt das in der Leber bereits abgelagerte Eisen ab. Durch die langsame Kinetik der Eisenaufnahme und -abgabe aus der Leber sind Änderungen der LIC insbesondere im langfristigen Verlauf zu erwarten. Die LIC wird mittels MRT gemessen und wird entsprechend als Eisenkonzentration (mg/g) bezogen auf das Trockengewicht (*dry weight*, d. w.) der Leber angegeben. Höhere Werte entsprechen dabei einem höheren Risiko für Siderose-bedingte Leberschäden.

Betrachtet wird im Dossier die mittlere Änderung der LIC zu Woche 48 (und ergänzend zu Woche 24) gegenüber Baseline. Die Auswertung erfolgte analog zum sFerritin mittels ANCOVA. Als Baseline-Wert wurde die letzte Messung vor Behandlungsbeginn verwendet. Die Ergebnisse finden sich in Anhang 4-G.

### *Änderung des Eisenchelator-Bedarfs*

Dosisanpassungen von Eisenchelatoren sollen gemäß Zulassung auf Basis der gemessenen Eisenparameter alle 2 bis 6 Monate erfolgen [37, 38]. Änderungen des Eisenchelator-Bedarfs sind dementsprechend vor allem bei langfristiger Beobachtung zu erwarten.

Im Dossier wird die mittlere Änderung der durchschnittlichen täglichen Eisenchelator-Dosis pro Wirkstoff in den letzten 24 Wochen der Behandlungsphase bis Woche 48 gegenüber den 24 Wochen vor Behandlungsbeginn dargestellt. Die Auswertung erfolgte jeweils analog zum sFerritin mittels ANCOVA, wobei alle Patient:innen berücksichtigt wurden, die eine Eisenchelator-Monotherapie erhielten und ihren Wirkstoff im relevanten Zeitraum nicht wechselten. Die Ergebnisse sind in Anhang 4-G zu finden.

## **Patientenberichtete Endpunkte zur Symptomatik und Lebensqualität**

Zusätzlich zu den Hauptanalysen (Zeit bis zur ersten Verbesserung / Zeit bis zur ersten Verschlechterung) werden für patientenberichtete Endpunkte weitere Operationalisierungen als unterstützende Evidenz betrachtet. Diese umfassen jeweils Responderanalysen zum Anteil der Patient:innen mit einer Verbesserung bzw. Verschlechterung, MMRM-Analysen zur Auswertung der stetigen Änderung bis Woche 48 gegenüber dem Ausgangswert (für alle patientenberichteten Skalen mit Ausnahme des PGI-C), sowie graphische Darstellungen der Veränderungen über den gesamten Studienverlauf.

### *Responderanalysen*

Im Rahmen von Responderanalysen wird der Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung sowie einer bewertungsrelevanten Verschlechterung zu Woche 48 (und ergänzend zu Woche 24) dargestellt. Patient:innen mit einem fehlenden Wert zu Woche 48 (bzw. Woche 24) wurden von diesen Analysen ausgeschlossen. Als Schätzer des Behandlungseffekts wird das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.

### *MMRM-Analyse*

Für die Auswertung der Veränderung gegenüber dem Ausgangswert als stetige Variable wurde jeweils ein MMRM auf Basis der Daten aus dem Beobachtungszeitraum bis Woche 48 herangezogen (für alle patientenberichteten Skalen mit Ausnahme des PGI-C). Im entsprechenden Modell wurden, zusätzlich zur Behandlungsgruppe, der jeweilige Baseline-Score, die Visite, die Interaktion Behandlungsgruppe\*Visite, die Interaktion Baseline-Score\*Visite sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren als feste Effekte verwendet. Der Achsenschnittpunkt (Intercept) und die Visite wurden zudem als zufällige Effekte definiert. Zur Quantifizierung des Behandlungseffekts wird der adjustierte Modellschätzer der Mittelwertdifferenz zu Woche 48 (und ergänzend zu Woche 24) sowie die hieraus abgeleitete standardisierte Mittelwertdifferenz (SMD) in Form von Hedges' g darstellt.

### *Graphische Darstellung der mittleren Änderung*

Um die mittlere Änderung auch im zeitlichen Verlauf abzubilden, werden die arithmetischen Mittelwerte pro Behandlungsarm und Visite im Rahmen von Line-Plots dargestellt. Für den PGI-C (bei dem die Veränderung nicht gegenüber dem Ausgangswert berechnet wurde, sondern direkt durch die Patient:innen auf einer kategoriellen 7-Punkte-Skala bewertet wurde) wird anstelle des Line-Plots ein Balkendiagramm mit der Verteilung der Bewertungen pro Visite verwendet.

### *Ereigniszeitanalysen unter Verwendung eines alternativen Schwellenwerts (PGI-C)*

Die Hauptanalyse bzgl. der Symptomatik gemäß PGI-C betrachtet die beiden besten Bewertungen der Veränderung („viel besser“ und „sehr viel besser“) als bewertungsrelevante Verbesserung und die beiden schlechtesten Bewertungen („viel schlechter“ und „sehr viel schlechter“) als bewertungsrelevante Verschlechterung. Als zusätzliche Analyse wird eine Auswertung des PGI-C bzgl. jeglicher Verbesserung/Verschlechterung der Symptomatik betrachtet (d. h. als Verbesserung zählen die Antworten „ein wenig besser“, „viel besser“ und „sehr viel besser“, als Verschlechterung die Antworten „ein wenig schlechter“, „viel schlechter“ und „sehr viel schlechter“). Somit können auch geringe Verbesserungen/Verschlechterungen der Symptomatik in die Bewertung einfließen.

## **Sicherheit**

Zusätzlich zu den UE-Hauptkategorien sowie den geforderten Analysen nach SOC/PT werden im Dossier UE von besonderem Interesse und weitere untersuchte UE und Hospitalisierungen dargestellt.

### *UE von besonderem Interesse und weitere untersuchte UE*

Die UE von besonderem Interesse (AESI) und weitere untersuchte UE umfassen in der Studie BEYOND die folgenden Kategorien:

- UE von besonderem Interesse
  - Malignitäten

- Prämaligne Erkrankungen
- Weitere untersuchte UE
  - Thromboembolische Ereignisse und Thrombophlebitis
  - Knochenschmerzen und knochenschmerzähnliche Ereignisse
  - Hypertonie
  - Extramedulläre Hämatopoese
  - Traumatische Frakturen
  - Beinulcera

Luspatercept greift als Erythrozyten-Reifungs-Aktivator (ERA) in die Erythropoese ein. Um auszuschließen, dass dieser Eingriff in die Hämatopoese bösartige Aberrationen zur Folge hat, wurden Malignitäten (definiert durch die *Standardised MedDRA Query* [SMQ] „Malignitäten“ unter Ausschluss der PTs „Humanes Choriongonadotropin positiv“, „Humanes Choriongonadotropin erhöht“, „Leukostasesyndrom“, „Tumorlysesyndrom“, „Gammopathie“ und „Monoklonale Gammopathie“) und prämaligne Erkrankungen (definiert durch die SMQ „Prämaligne Erkrankungen“) jeweils als UE von besonderem Interesse betrachtet.

Weiterhin wurden Thromboseereignisse und Thrombophlebitis (definiert durch die SMQ „Embolie- und Thromboseereignisse“ sowie den breiten Umfang der SMQ „Thrombophlebitis“ unter Ausschluss der PTs „Verschluss eines medizinischen Gerätes“, „Durch Medizinprodukt bedingte Thrombose“, „Thrombose in einem Medizinprodukt“ und „Transitorische ischämische Attacke“), Knochenschmerzen und knochenschmerzähnliche Ereignisse (ad hoc zusammengesetzt aus den PTs „Rückenschmerzen“, „Knochenschmerzen“, „Schmerzen des Muskel- und Skelettsystems“, „Nackenschmerzen“, „Schmerz in einer Extremität“, „Kieferschmerzen“ und „Wirbelsäulenschmerz“) und Hypertonie (definiert durch den schmalen Umfang der SMQ „Hypertonie“) als gesonderte UE-Kategorien (*targeted/tracked safety events*) betrachtet. Als weitere relevante UE werden zudem, analog zum CSR, traumatische Frakturen (entsprechend dem PT „Traumatische Fraktur“) und im Zusammenhang mit einer extramedullären Hämatopoese stehende Ereignisse (entsprechend den PTs „Extramedulläre Hämatopoese“ und „Raumforderung in der Brustwand“) angeführt. Darüber hinaus werden Beinulcera im Dossier als zusätzliche patientenrelevante UE-Kategorie berichtet (basierend auf allen identifizierten Ereignissen mit relevanten MedDRA *Lowest Level Terms*: „Knöchel Ulkus“, „Fußulkus“, „Infiziertes Hautulkus“, „Offenes Bein“).

Die Auswertung im Dossier erfolgte analog zu den UE-Hauptkategorien als Ereigniszeitanalyse. Zusätzlich zur Gesamtbetrachtung der Kategorien, werden schwere (CTCAE-Grad  $\geq 3$ ) und schwerwiegende Ereignisse separat dargestellt.

### *Hospitalisierungen*

Als Hospitalisierungen gezählt wurden unabhängig von ihrer Ursache alle dokumentierten stationären Krankenhausaufenthalte während der Studie. Die Auswertung erfolgte anhand einer Ereigniszeitanalyse auf Basis der Zeit von der ersten Dosis bis zur ersten Hospitalisierung. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne Hospitalisierung wurden zum Zeitpunkt der Follow-up-Visite (ca. 9 Wochen nach Behandlungsende) bzw. zum Zeitpunkt des betrachteten Datenschnitts oder bei vorherigem Studienabbruch/Tod, je nachdem welches Ereignis früher eintrat, zensiert. Um eine bessere Beurteilung der Relevanz des Endpunkts Hospitalisierungen für die Nutzenbewertung zu ermöglichen, werden zudem die Gründe der Hospitalisierungen dargestellt. Hierzu wurde die Häufigkeit der unterschiedlichen Gründe anhand des Anteils betroffener Patient:innen pro Studienarm ausgewertet. Da UE, die zu Hospitalisierungen führten, bereits bei der Auswertung der schwerwiegenden UE berücksichtigt werden, wird des Weiteren eine Auswertung der Hospitalisierungen vorgenommen, ohne solche Hospitalisierungen zu berücksichtigen, die durch UE bedingt wurden, um eine doppelte Berücksichtigung dieser Ereignisse zu vermeiden.

#### **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.*

Gemäß dem statistischen Analyseplan der Studie BEYOND waren Subgruppenanalysen hinsichtlich ausgewählter Merkmale u. a. für den primären und sekundären Endpunkt festgelegt. Im vorliegenden Dossier wurden Subgruppenanalysen unter Berücksichtigung dieser präspezifizierten Merkmale für sämtliche patientenrelevante Endpunkte vorgenommen. Die betrachteten Subgruppenmerkmale und Subgruppen sind in der nachfolgenden Tabelle 4-4 aufgeführt. Die Trennpunkte der Subgruppen waren a priori im statistischen Analyseplan definiert.

Tabelle 4-4: Subgruppenanalysen der Studie BEYOND

Subgruppenmerkmal	Subgruppe
Geschlecht	Weiblich Männlich
Alter	≤ 32 Jahre > 32 Jahre
Region	Nordamerika/Europa Asien-Pazifik Mittlerer Osten
Beta-Thalassämie Diagnose	Beta-Thalassämie Hämoglobin E/Beta-Thalassämie Beta-Thalassämie kombiniert mit Alpha-Thalassämie
Vorhergehende Splenektomie	Ja Nein
Hb-Wert zu Baseline	≥ 8,5 g/dl < 8,5 g/dl
Lebereisenkonzentration (LIC) zu Baseline, gemessen mittels MRT	< 3 mg/g Trockengewicht (d. w.) 3 – 5 mg/g d. w. > 5 – 7 mg/g d. w. > 7 – 15 mg/g d. w. > 15 mg/g d. w.

Subgruppenmerkmal	Subgruppe
NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche zu Baseline	≥ 3 Punkte < 3 Punkte
6-Minuten-Gehtest zu Baseline	≤ 450 m > 450 m
FACIT-Fatigue Skala zu Baseline	≥ 43 Punkte < 43 Punkte
d. w.: Trockengewicht; FACIT: <i>Functional Assessment of Chronic Illness Therapy</i> ; Hb: Hämoglobin; MRT: Magnetresonanztomographie; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i>	

### Methodik von Subgruppenanalysen

Zur Identifikation potenzieller Effektmodifikatoren wurde pro Endpunkt ein Interaktionstest für jedes Subgruppenmerkmal berechnet. Hierfür wurde für Ereigniszeitanalysen der Interaktionsterm aus einem unstratifizierten Cox-Regressionsmodell mit Behandlungsarm, Subgruppenmerkmal und Behandlungsarm\*Subgruppenmerkmal sowie ggf. dem Baseline-Wert herangezogen. Analog hierzu wurde der Interaktionsterm bei stetigen Veränderungen basierend auf einer ANCOVA mit den unabhängigen Variablen Behandlungsarm, Baseline-Wert, Subgruppenmerkmal und Behandlungsarm\*Subgruppenmerkmal bestimmt.

Subgruppen mit weniger als 10 Patient:innen (bezogen auf beide betrachteten Studienarme zusammengenommen) werden bei der Durchführung der Subgruppenanalysen nicht berücksichtigt. Bei Ereigniszeitanalysen wurde darüber hinaus für die Subgruppenanalyse hinsichtlich eines bestimmten Merkmals nur dann berichtet, wenn für den jeweiligen Endpunkt in mindestens einer der entsprechenden Subgruppen mindestens 10 Ereignisse auftraten (vgl. Angaben in Abschnitt 4.3.1.3.2).

### Interpretation der Ergebnisse der Subgruppenanalysen

Bei der Bewertung der Subgruppenergebnisse sind die generellen Limitationen der Aussagekraft dieser Analysen zu beachten. Im Allgemeinen sind Subgruppenergebnisse als hypothesengenerierend und nicht als konfirmatorische Überprüfung einer Hypothese zu betrachten. Zudem sind die Ergebnisse der Interaktionstests insbesondere dann sehr vorsichtig zu interpretieren, wenn die Patientenzahlen in den Subgruppen ungleich bzw. unverhältnismäßig zwischen den Behandlungsarmen verteilt sind oder wenn eine bestimmte Subgruppe in einem Behandlungsarm nur sehr wenige Patient:innen umfasst. Die Ableitung separater Nutzensaussagen für Subgruppen allein auf Basis signifikanter Interaktionstests wäre daher mit hoher Unsicherheit behaftet und wird von BMS nicht als adäquat betrachtet. Neben dem Interaktionstest sollte immer auch die medizinische oder (patho-)physiologische Plausibilität der Subgruppenergebnisse geprüft und in der Bewertung berücksichtigt werden.

#### 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 darstellen<sup>9</sup>. Eine Ausnahme kann das Vorliegen von dramatischen Effekten sein. An Stelle von nicht adjustierten indirekten Vergleichen sollen je nach Datenlage einfache adjustierte indirekte Vergleiche<sup>10</sup> 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)<sup>11</sup> und Rücker (2012)<sup>12</sup> 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 ist<sup>13</sup>.

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<sup>14, 15, 16</sup>

---

<sup>9</sup> 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](http://www.gmds.de/pdf/publikationen/stellungnahmen/120202_IQWIG_GMDS_IBS_DR.pdf).

<sup>10</sup> 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.

<sup>11</sup> Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.

<sup>12</sup> Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012; 3(4): 312-324.

<sup>13</sup> Schöttker B, Lüthmann D, Boukhemair D, Raspe H. Indirekte Vergleiche von Therapieverfahren. Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.

<sup>14</sup> 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.

<sup>15</sup> 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

<sup>16</sup> 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.*
- *Genauere 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.*

Ein indirekter Vergleich wurde nicht durchgeführt.

### 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 4-5: 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
BEYOND	ja	ja	abgeschlossen	03/2018–11/2022  <u>Primärer Datenschnitt:</u> 14.09.2020 (präspezifizierte Analyse nach Beendigung der doppelblinden Behandlungsphase) <u>Ad-hoc Follow-up-Datenschnitte:</u> 15.03.2021 und 22.09.2021 (Unterstützung des Zulassungsverfahrens) <u>Finale Daten:</u> 14.02.2023 (Datenbankschluss nach Beendigung der Studie)	a) Luspatercept + BSC b) Placebo + BSC
BSC: Best-Supportive-Care					

Geben Sie an, welchen Stand die Information in Tabelle 4-5 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 Information: 21.02.2023.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-5 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-6: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
/	Nicht zutreffend.

#### 4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche

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

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

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

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

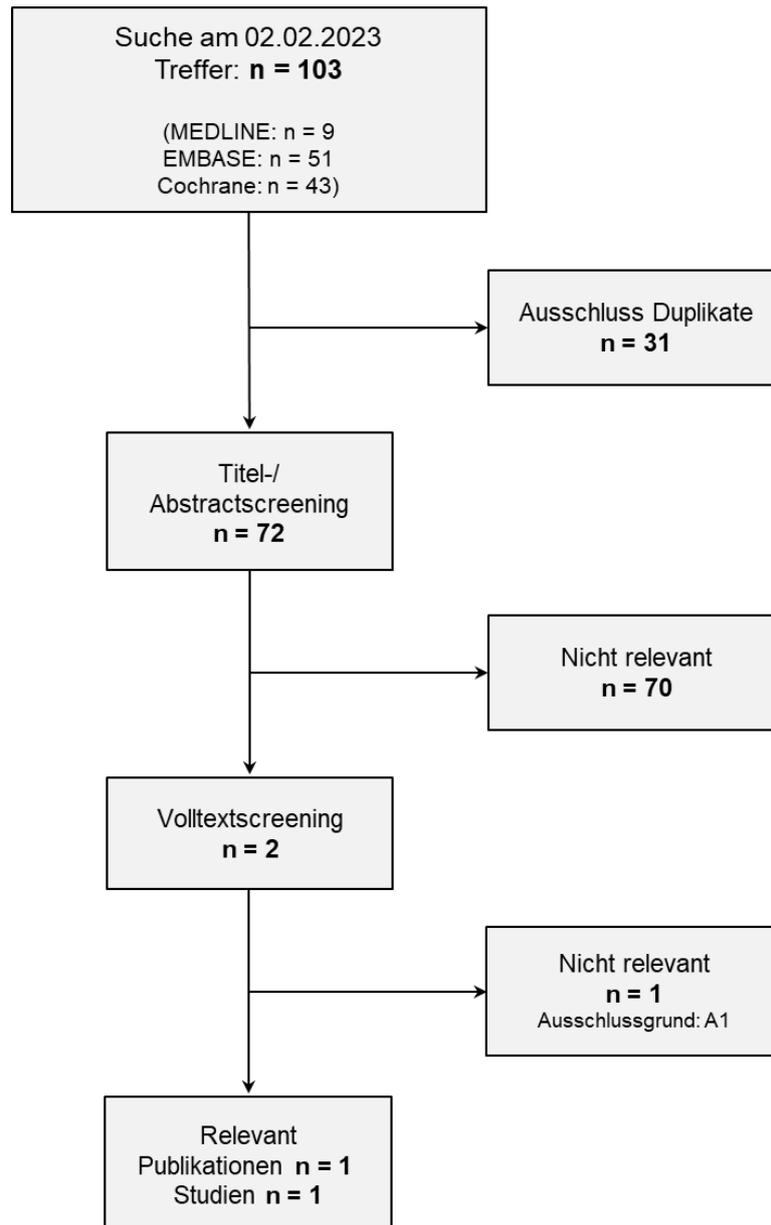


Abbildung 4-1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel

Am 02.02.2023 wurde eine bibliografische Literaturrecherche in den Datenbanken MEDLINE, EMBASE und The Cochrane Library zur Identifizierung von relevanten RCT mit dem zu bewertenden Arzneimittel Luspatercept durchgeführt. Zur Bewertung der Relevanz wurden die in Abschnitt 4.2.2 definierten Kriterien für den Ein- und Ausschluss von Studien herangezogen. Insgesamt ergab die Recherche 103 Treffer, von denen 72 nach Entfernung der Dubletten (n = 31) anhand der Titel und Abstracts gesichtet wurden. Publikationen, deren Relevanz anhand der Titel und/oder Abstracts nicht eindeutig erkennbar war, wurden anschließend anhand der Volltexte bewertet. Nicht relevante Publikationen wurden ausgeschlossen (n = 70 nach Titel-/Abstractscreening, n = 1 nach Volltextscreening), sodass eine relevante Publikation verblieb, welche sich auf die relevante Studie BEYOND bezieht.

#### 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 4-5) und ob die Studie auch durch die bibliografische Literaturrecherche identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein. Listen Sie die ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-D.

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

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

Tabelle 4-7: 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 <sup>a</sup> )	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)	Status (abgeschlossen/ abgebrochen/ laufend)
BEYOND	<a href="https://clinicaltrials.gov/ct2/show/study/NCT03342404">clinicaltrials.gov: NCT03342404</a> [Studienregistereintrag] [39] <a href="https://www.ictrp.org/ct2/show/study/NCT03342404">ICTRP: NCT03342404</a> [Studienregistereintrag] [40]	ja	ja [41]	abgeschlossen

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.

Geben Sie an, welchen Stand die Information in Tabelle 4-7 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.

Stand der Information: 02.02.2023.

#### 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 4-5) 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 4-8: Relevante Studien aus der Suche auf der Internetseite des G-BA – RCT mit dem zu bewertenden Arzneimittel

Studie	Relevante Quellen <sup>a</sup>	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)
-	-	-	-	-

a: Quellen aus der Suche auf der Internetseite des G-BA

Geben Sie an, welchen Stand die Information in Tabelle 4-8 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.

Stand der Information: 06.02.2023.

#### 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 4-9: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	Studienkategorie			verfügbare Quellen <sup>a</sup>		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie <sup>b</sup> (ja/nein)	Studie Dritter (ja/nein)	Studienberichte (ja/nein [Zitat])	Register-einträge <sup>c</sup> (ja/nein [Zitat])	Publikation und sonstige Quellen <sup>d</sup> (ja/nein [Zitat])
<b>Placebokontrolliert + BSC<sup>e</sup></b>						
BEYOND	ja	ja	nein	ja [42]	ja [39, 40]	ja [41]
<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: BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie</p> <p>BSC: Best-Supportive-Care</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 4-10: 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
BEYOND	RCT, doppelblind, parallel	Erwachsene, Anämie aufgrund von nicht-transfusionsabhängiger Beta-Thalassämie	Luspatercept + BSC <sup>(1)</sup> (n = 96) Placebo + BSC <sup>(1)</sup> (n = 49)	<u>Screening:</u> 4 Wochen <u>Doppelblinde Behandlungsphase:</u> Min. 48 Wochen <sup>(2)</sup> <u>Offene Behandlungsphase:</u> Ab Entblindung Luspatercept in beiden Behandlungsarmen <sup>(3)</sup> <u>Nachbeobachtung:</u> 3 Jahre nach letzter Gabe der Studienmedikation  <u>Primärer Datenschnitt:</u> 14.09.2020 (präspezifizierte Analyse nach Beendigung der doppelblinden Behandlungsphase) <u>Ad-hoc Follow-up-Datenschnitte:</u> 15.03.2021 und 22.09.2021 (Unterstützung des Zulassungsverfahrens) <u>Finale Daten:</u> 14.02.2023 (Datenbank-schluss nach Beendigung der Studie)	<u>Regionen:</u> Nordamerika/ Europa (Griechenland, Italien, Vereinigtes Königreich, USA), Mittlerer Osten (Libanon), Asien-Pazifik (Thailand)  <u>Zeitraum:</u> 03/2018-11/2022 <sup>(4)</sup>	<u>Primärer Endpunkt:</u> Anteil der Patient:innen mit einer Erhöhung der mittleren Hämoglobin-Konzentration im Vergleich zu Baseline um $\geq 1,0$ g/dl ohne Transfusion über einen kontinuierlichen Zeitraum von zwölf Wochen (Woche 13 bis 24) <u>Patientenrelevante sekundäre Endpunkte:</u> NTDT-PRO, Transfusionsfreiheit, 6MWT, SF-36, FACIT-Fatigue Skala, unerwünschte Ereignisse
<p><sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.  <sup>(2)</sup> Die Behandlung in der Studie wurde doppelblind durchgeführt, bis alle Patient:innen für 48 Wochen behandelt worden waren oder die Behandlung vorzeitig abgebrochen hatten.  <sup>(3)</sup> Mit Abschluss der 48-wöchigen doppelblinden Behandlungsphase durch den letzten Patienten bzw. die letzte Patientin wurde die Studie entblindet. Nach der Entblindung konnten die Patient:innen beider Studienarme im Rahmen einer offenen Behandlungshase (<i>Open-Label-Treatment-Period</i>, OLP) für einen Zeitraum von ca. 24 Monaten mit Luspatercept behandelt werden (Voraussetzung für die Umstellung von Patient:innen aus dem Placebo-Arm auf Luspatercept war eine entsprechende Empfehlung des DMC). Anschließend konnten Patient:innen in die <i>Roll-Over</i>-Studie ACE-536-LTFU-001 wechseln, um weiterhin Luspatercept zu erhalten.  <sup>(4)</sup> Erste Visite der ersten Patientin bzw. des ersten Patienten bis letzte Visite der letzten Patientin bzw. des letzten Patienten.          6MWT: 6-Minuten-Gehtest; BSC: <i>Best-Supportive-Care</i>; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy-Fatigue</i>; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i>; RCT: randomisierte kontrollierte Studie (<i>randomized controlled trial</i>); SF-36 = <i>Short Form (36) Health Survey</i>; USA: Vereinigte Staaten von Amerika</p>						

Tabelle 4-11: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	Luspatercept + BSC <sup>(1)</sup>	Placebo + BSC <sup>(1)</sup>	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in-Phase etc.</i>
BEYOND	0,45 – 1,25 mg/kg Körpergewicht Luspatercept s. c. alle drei Wochen	Volumenäquivalent Placebo zur Luspatercept- Gruppe s. c. alle drei Wochen	<u>Dosistitration</u> bis maximal 1,25 mg/kg Körpergewicht war unter Berücksichti- gung der Hb-Konzentration erlaubt. Die maximale Gesamtdosis von 120 mg durfte nicht überschritten werden.  In beiden Behandlungsarmen war eine supportive Behandlung mit Antibiotika, Antimykotika, Virostatika, und/oder unterstützenden Ernährungsmaßnahmen, sofern benötigt, erlaubt <sup>(2)</sup>
<p><sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.</p> <p><sup>(2)</sup> Während der Studie explizit ausgeschlossen waren die Behandlung mit anderen Prüfsubstanzen, Hydroxyurea, Anagrelid und hämatopoetischen Wachstumsfaktoren. Sofern eine Behandlung mit Hydroxyurea während der doppelblinden Behandlungsphase (Woche 1 – 48) notwendig war, wurde die Behandlung mit der Studienmedikation abgebrochen und der Patient bzw. die Patientin ging in die Nachbeobachtungsphase über.</p> <p>BSC: <i>Best-Supportive-Care</i>; Hb: Hämoglobin; s. c.: subkutan</p>			

Tabelle 4-12: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N = 96)	Placebo + BSC <sup>(1)</sup> (N = 49)	Gesamt (N = 145)
<b>Demographische Charakteristika zu Baseline</b>			
Alter [Jahre]			
Mittelwert (SD)	39,3 (13,2)	41,1 (11,9)	39,9 (12,8)
Median (Min; Max)	39,5 (18; 71)	41,0 (19; 66)	40,0 (18; 71)
Altersgruppe [n (%)]			
≤ 32 Jahre	35 (36,5)	12 (24,5)	47 (32,4)
> 32 – ≤ 50 Jahre	41 (42,7)	26 (53,1)	67 (46,2)
> 50 Jahre	20 (20,8)	11 (22,4)	31 (21,4)
Geschlecht [n (%)]			
männlich	40 (41,7)	23 (46,9)	63 (43,4)
weiblich	56 (58,3)	26 (53,1)	82 (56,6)
Abstammung [n (%)]			
asiatisch	31 (32,3)	13 (26,5)	44 (30,3)
weiß	59 (61,5)	28 (57,1)	87 (60,0)
andere	6 (6,3)	8 (16,3)	14 (9,7)
BMI [kg/m <sup>2</sup> ]			
Mittelwert (SD)	22,1 (3,8)	22,3 (4,1)	22,2 (3,9)
Median (Min; Max)	21,6 (15;37)	21,5 (16;38)	21,6 (15;38)

<b>Studie BEYOND</b>	<b>Luspatercept + BSC <sup>(1)</sup></b> <b>(N = 96)</b>	<b>Placebo + BSC <sup>(1)</sup></b> <b>(N = 49)</b>	<b>Gesamt</b> <b>(N = 145)</b>
<b>BMI [n (%)]</b>			
< 20 kg/m <sup>2</sup>	35 (36,5)	18 (36,7)	53 (36,6)
20 – < 25 kg/m <sup>2</sup>	46 (47,9)	20 (40,8)	66 (45,5)
≥ 25 – < 30 kg/m <sup>2</sup>	12 (12,5)	9 (18,4)	21 (14,5)
≥ 30 kg/m <sup>2</sup>	3 (3,1)	2 (4,1)	5 (3,4)
<b>Region [n (%)]</b>			
Nordamerika und Europa	60 (62,5)	30 (61,2)	90 (62,1)
Mittlerer Osten	9 (9,4)	8 (16,3)	17 (11,7)
Asien/Pazifik	27 (28,1)	11 (22,4)	38 (26,2)
<b>Krankheitsspezifische Charakteristika</b>			
<b>Beta-Thalassämie-Diagnose [n (%)]</b>			
Beta-Thalassämie	63 (65,6)	34 (69,4)	97 (66,9)
HbE/Beta-Thalassämie	28 (29,2)	11 (22,4)	39 (26,9)
Beta-Thalassämie kombiniert mit Alpha-Thalassämie	5 (5,2)	4 (8,2)	9 (6,2)
<b>Beta-Thalassämie-Genotyp [n (%)]</b>			
β0/β0, β+/β+, β+/β0 ohne Alpha-Thalassämie	69 (71,9)	33 (67,3)	102 (70,3)
β0/β0, β+/β+, β+/β0 mit Alpha-Thalassämie	6 (6,3)	4 (8,2)	10 (6,9)
β0/β, β+/β mit Alpha-Gen- Duplikation	21 (21,9)	12 (24,5)	33 (22,8)
<b>Hb-Wert zu Baseline [g/dl] <sup>(2)</sup></b>			
Mittelwert (SD)	8,2 (1,22)	8,1 (1,30)	8,2 (1,24)
Median (Min; Max)	8,2 (5,3;10,1)	8,1 (5,7; 10,1)	8,2 (5,3; 10,1)
<b>[n (%)]</b>			
≥ 8,5 g/dl	41 (42,7)	20 (40,8)	61 (42,1)
< 8,5 g/dl	55 (57,3)	29 (59,2)	84 (57,9)
<b>NTDT-PRO-Gesamtscore Müdigkeit/Schwäche zu Baseline <sup>(3)</sup></b>			
Mittelwert (SD)	4,1 (2,29)	4,00 (2,03)	4,0 (2,20)
Median (Min; Max)	4,3 (0,0; 9,5)	4,1 (0,4; 9,5)	4,3 (0,0; 9,5)
<b>[n (%)]</b>			
≥ 3 Punkte	66 (68,8)	35 (71,4)	101 (69,7)
< 3 Punkte	30 (31,3)	14 (28,6)	44 (30,3)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N = 96)	Placebo + BSC <sup>(1)</sup> (N = 49)	Gesamt (N = 145)
Transfusionslast vor der ersten Dosis [EK-Einheiten/24 Wochen]			
Mittelwert (SD)	0,3 (0,8)	0,3 (0,8)	0,3 (0,8)
Median (Min; Max)	0,0 (0; 4)	0,0 (0; 4)	0,0 (0; 4)
[n (%)]			
0 EK-Einheiten/24 Wochen	83 (86,5)	42 (85,7)	125 (86,2)
> 0 EK-Einheiten/24 Wochen	13 (13,5)	7 (14,3)	20 (13,8)
Vorhergehende Splenektomie [n (%)]			
ja	34 (35,4)	26 (53,1)	60 (41,4)
nein	62 (64,6)	23 (46,9)	85 (58,6)
Lebereisenkonzentration (LIC) zu Baseline (mittels MRT) <sup>(4)</sup> [mg/g d. w.]			
(n = 95) (n = 47) (n = 142)			
Mittelwert (SD)	6,1 (6,15)	5,9 (5,82)	6,0 (6,02)
Median (Min; Max)	3,9 (0,8; 39,9)	4,1 (0,7; 28,7)	4,0 (0,7; 39,9)
[n (%)]			
< 3 mg/g d. w.	35 (36,5)	17 (34,7)	52 (35,9)
≥ 3 – ≤ 5 mg/g d. w.	23 (24,0)	13 (26,5)	36 (24,8)
> 5 – ≤ 7 mg/g d. w.	12 (12,5)	3 (6,1)	15 (10,3)
> 7 – ≤ 15 mg/g d. w.	14 (14,6)	9 (18,4)	23 (15,9)
> 15 mg/g d. w.	11 (11,5)	5 (10,2)	16 (11,0)
fehlend	1 (1,0)	2 (4,1)	3 (2,1)
Serum-Ferritin-Wert zu Baseline <sup>(5)</sup> [µg/l]			
Mittelwert (SD)	567,8 (523,16)	528,8 (444,92)	554,6 (496,9)
Median (Min; Max)	456,5 (30,0; 3528,0)	360,0 (40,0; 2265,0)	441,0 (30,0; 3528,0)
[n (%)]			
< 500 µg/l	55 (57,3)	28 (57,1)	83 (57,2)
≥ 500 – ≤ 800 µg/l	28 (29,2)	9 (18,4)	37 (25,5)
> 800 µg/l	13 (13,5)	12 (24,5)	25 (17,2)
Eisenchelatherapie vor Studienbeginn (in den 24 Wochen vor Studienbeginn) [n (%)]			
ja	28 (29,2)	16 (32,7)	44 (30,3)
nein	68 (70,8)	33 (67,3)	101 (69,7)
ECOG Performance-Status zu Baseline [n (%)]			
0	62 (64,6)	38 (77,6)	100 (69,0)
1	34 (35,4)	11 (22,4)	45 (31,0)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N = 96)	Placebo + BSC <sup>(1)</sup> (N = 49)	Gesamt (N = 145)
6-Minuten-Gehtest-Messung zu Baseline [m]			
Mittelwert (SD)	447,0 (93,0)	435,0 (78,3)	442,9 (88,2)
Median (Min; Max)	447,7 (0; 661)	430,0 (300; 600)	443,0 (0; 661)
[n (%)]			
≤ 450 m	51 (53,1)	31 (63,3)	82 (56,6)
> 450 m	45 (46,9)	18 (36,7)	63 (43,4)
FACIT-Fatigue-Score zu Baseline [n (%)]			
≥ 43 Punkte	32 (33,3)	16 (32,7)	48 (33,1)
< 43 Punkte	64 (66,7)	32 (65,3)	96 (66,2)
fehlend	0	1 (2,0)	1 (0,7)
Population: ITT-Population; primärer Datenschnitt vom 14.09.2020			
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.			
<sup>(2)</sup> Mittlerer Hb-Wert aus mindestens 2 dokumentierten Hb-Werten innerhalb der Screening-Phase.			
<sup>(3)</sup> Mittlerer NTDT-PRO-Score bezogen auf die 7 Tage vor der ersten Dosis			
<sup>(4)</sup> Der LIC-Wert basiert entweder auf dem elektronischen Patientenbericht oder dem abgeleiteten Wert vom T2*-, R2*- oder R2-Parameter, je nachdem welche Technik und Software zur Bestimmung der LIC-MRT benutzt wurde.			
<sup>(5)</sup> Mittlerer Serum-Ferritin-Wert bezogen auf die 24 Wochen vor der ersten Dosis			
BMI: <i>Body-Mass-Index</i> ; BSC: <i>Best-Supportive-Care</i> ; ECOG: <i>Eastern Cooperative Oncology Group</i> ; EK: Erythrozytenkonzentrat; FACIT: <i>Functional Assessment of Chronic Illness Therapy</i> ; Hb: Hämoglobin; HbE: Hämoglobin E; ITT: <i>Intention-to-Treat</i> ; Max: Maximum; Min: Minimum; MRT: <i>Magnetresonanztomographie</i> ; RNA: <i>Ribonukleinsäure</i> ; SD: Standardabweichung			

*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.*

### Studie BEYOND

Die Studie BEYOND ist eine randomisierte kontrollierte Doppelblindstudie, in der die Wirksamkeit und Sicherheit von Luspatercept im Vergleich zu Placebo, jeweils ergänzt durch eine bestmöglich unterstützende Therapie zur Symptomkontrolle und Verbesserung der Lebensqualität (*Best-Supportive-Care*), bei Patient:innen mit einer nicht-transfusionsabhängigen Beta-

Thalassämie (NTDT) untersucht wurde. Geeignete Patient:innen wurden im Verhältnis 2:1 in den Luspatercept-Arm oder Placebo-Arm randomisiert, wobei eine Stratifizierung nach Hämoglobin-Konzentration ( $\geq 8,5$  g/dl vs.  $< 8,5$  g/dl) und NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche zu Baseline ( $\geq 3$  Punkte vs.  $< 3$  Punkte) erfolgte.

Das primäre Ziel der Studie war die Untersuchung der Wirksamkeit von Luspatercept, gemessen anhand der durchschnittlichen Hämoglobin-Konzentration ohne Transfusionen über einen Zeitraum von 12 Wochen, zwischen Woche 13 und 24, im Vergleich zur Baseline.

### ***Patientenpopulation***

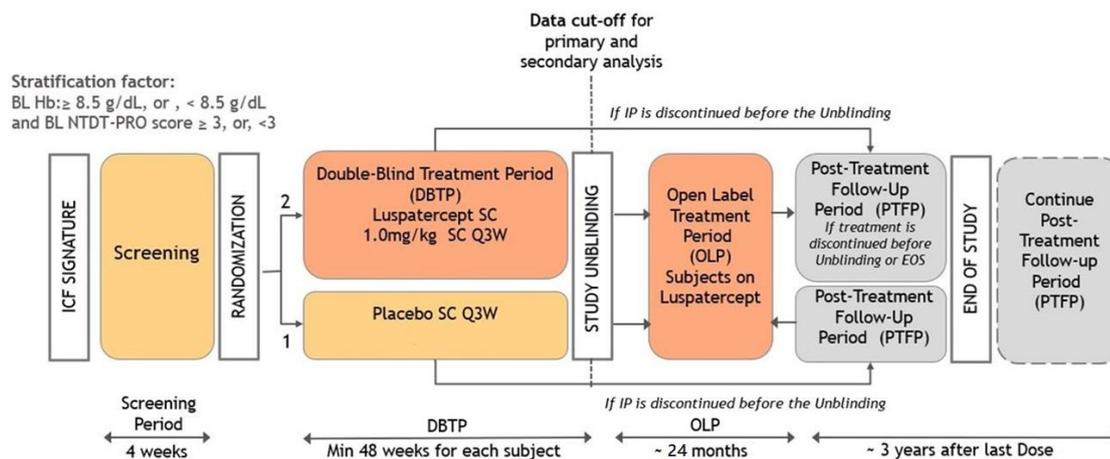
Die einzuschließende Studienpopulation umfasste erwachsene Patient:innen mit einer dokumentierten Beta-Thalassämie oder HbE/Beta-Thalassämie, sowie einem mittleren Hb  $\leq 10$  g/dl (bezogen auf die letzten 4 Wochen vor der Randomisierung). Darüber hinaus durften Patient:innen in den 24 Wochen vor Randomisierung maximal 5 EK-Einheiten erhalten haben und mussten in den 8 Wochen vor der Randomisierung gänzlich transfusionsfrei gewesen sein.

Durch die Einschlusskriterien wurde erstens sichergestellt, dass die Patient:innen symptomatisch und damit behandlungsbedürftig waren. So geht ein Hb  $\leq 10$  g/dl mit einer unzureichenden Sauerstoffzufuhr des Gewebes einher und das Risiko für Anämie-assoziierte Symptome ist ab dieser Schwelle stark erhöht. Dementsprechend zeigen Patient:innen mit einer NTDT bei einem Hb  $\geq 10$  g/dl ein signifikant längeres morbiditätsfreies Überleben als Patient:innen mit niedrigeren Hb-Werten [43, 44]. Auch internationale Leitlinien und Quellen besagen, dass bei Hb-Werten  $< 10$  g/dl EK-Transfusionen angezeigt sein können [33, 35]. Darüber hinaus bestand durch das Kriterium von maximal 5 EK-Einheiten eine klare Abgrenzung zur transfusionsabhängigen Beta-Thalassämie (TDT). So waren in der pivotalen Studie BELIEVE, die für die frühe Nutzenbewertung für die Behandlung transfusionsabhängiger Beta-Thalassämie mit Luspatercept herangezogen wurde,  $\geq 6$  EK-Transfusionen in den letzten 24 Wochen als Transfusionsabhängigkeit definiert [45].

Ausgeschlossen von der Studie waren Patient:innen mit anomalen Strukturvarianten des Beta-Globin-Gens ohne Beta-Thalassämie-Phänotyp, wie der Sichelzellkrankheit (auch gemischt heterozygote Sichelzellkrankheit wie die HbS/Beta-Thalassämie entsprechend dem Ausschlusskriterium der ICD-10 D56.1 „Beta-Thalassämie“) oder Alpha-Thalassämien, da es sich hierbei um separat zu betrachtende Krankheitsbilder handelt, die nicht vom Anwendungsgebiet von Luspatercept umfasst sind [10, 46].

### ***Studiendesign und Therapie***

Die Studie BEYOND wurde an 12 Zentren in 6 Ländern durchgeführt und war unterteilt in eine 4-wöchige Screeningphase, eine doppelblinde Behandlungsphase (Woche 1 – min. 48), sowie eine sich anschließende offene Extensionsphase und/oder eine *Posttreatment-Follow-up*-Phase (siehe auch Abbildung 4-2).



ICF: Einwilligungserklärung (*informed consent form*); IP: Prüfpräparat (*investigational product*); Q3W: alle 3 Wochen; SC: subkutan (*subcutaneous*)

Abbildung 4-2: Übersicht des Studiendesigns der Studie BEYOND

### Screening-Phase

Während der initialen, 4-wöchigen Screeningphase wurde die Eignung der Patient:innen für die Studienteilnahme geprüft.

### Doppelblinde Behandlungsphase

Die Patient:innen wurden im Verhältnis 2:1 in den Luspatercept- bzw. Placebo-Arm randomisiert. Die geplante Fallzahl umfasste insgesamt ca. 150 Patient:innen.

Die erste Dosis der Studienmedikation wurde innerhalb von 3 Tagen nach Randomisierung verabreicht. Im Luspatercept-Arm wurde Luspatercept zu einer anfänglichen Dosierung von 1,0 mg/kg Körpergewicht alle 3 Wochen als subkutane Injektion verabreicht, welche in einen Dosisbereich von 0,45 mg/kg bis maximal 1,25 mg/kg Körpergewicht titriert werden durfte (genutzter Dosisbereich von 0,60 mg/kg bis maximal 1,25 mg/kg Körpergewicht). Dies entspricht der zugelassenen Dosierung von Luspatercept im vorliegenden Anwendungsgebiet gemäß Fachinformation [3]. Im Placebo-Arm erhielten Patient:innen eine subkutane Placebo-Injektion (Natriumchlorid-Lösung, 0,9 %), die analog zur Luspatercept-Behandlung alle 3 Wochen verabreicht wurde, um die Verblindung in der Studie aufrecht zu erhalten.

Während der gesamten Studiendauer waren Begleittherapien bedarfsgerecht in beiden Behandlungsarmen im Rahmen des BSC möglich. So konnte die Anämie bei niedrigen Hb-Werten bzw. bei Anämie-assoziierten Symptomen mit EK-Transfusionen im Ermessen des Prüfarztes behandelt werden. Des Weiteren war eine Behandlung mit Eisenchelatoren entsprechend dem Zulassungsstatus sowie den lokalen Leitlinien erlaubt. Dies entspricht dem Behandlungsstandard im vorliegenden Anwendungsgebiet.

Die doppelblinde Behandlungsphase endete, nachdem alle Patient:innen mindestens 48 Wochen behandelt worden waren oder die Therapie vorzeitig abgebrochen hatten. Zu diesem Zeitpunkt erfolgte die präspezifizierte Analyse des primären und der sekundären Endpunkte sowie die Entblindung der Studie.

#### *Offene Behandlungsphase*

Nach der Entblindung konnten sowohl die Patient:innen aus dem Luspatercept-Arm, die die doppelblinde Behandlungsphase ohne Therapieabbruch abgeschlossen hatten, als auch – basierend auf einer entsprechenden Empfehlung des DMC – die Patient:innen aus dem Placebo-Arm, die bis zu diesem Zeitpunkt an allen vorgesehenen Untersuchungen teilgenommen hatten (selbst wenn sie die Therapie vor Woche 48 abgebrochen hatten), im Rahmen der offenen Behandlungsphase (OLP) für ca. 24 Monate (weiterhin) eine Therapie mit Luspatercept erhalten und im Anschluss zur Weiterbehandlung in die Rollover-Studie ACE-536-LTFU-001 wechseln.

#### *Post-Treatment-Follow-up-Phase*

Patient:innen, die die Therapie in der Studie BEYOND abbrachen, wurden im Rahmen der Post-Treatment-Follow-up-Phase (bzw. nach Beendigung der Studie BEYOND im Rahmen der Roll-Over-Studie ACE-536-LTFU-001) für mindestens 5 Jahre nach der ersten Dosis und mindestens 3 Jahre nach der letzten Dosis nachbeobachtet.

#### *Datenschnitte*

Die Studie begann am 05. März 2018 mit der ersten Visite der ersten Patientin bzw. des ersten Patienten und wurde durch den Wechsel aller verbliebenen Studienteilnehmer:innen in die Roll-Over-Studie ACE-536-LTFU-001 mit der letzten Visite der letzten Patientin bzw. des letzten Patienten am 28. November 2022 beendet.

Die primäre Analyse war gemäß statistischem Analyseplan präspezifiziert auf den Zeitpunkt, zu dem alle Patient:innen die doppelblinde Behandlungsphase von mindestens 48 Wochen abgeschlossen oder abgebrochen hatten. Der Datenschnitt für die primäre Analyse erfolgte am 14. September 2020 (Datenbankschluss 23.11.2020). Neben dem primären Datenschnitt wurden zur Unterstützung des Zulassungsverfahrens am 15. März 2021 und am 22. September 2021 zwei weitere Follow-up-Datenschnitte vorgenommen (Datenbankschluss: 19.07.2021 bzw. 31.03.2022), die jedoch nicht präspezifiziert waren. Der finale Datenbankschluss nach Beendigung der Studie erfolgte am 14. Februar 2023.<sup>17</sup>

Für die Nutzenbewertung wird der primäre Datenschnitt auf Basis der doppelblinden Behandlungsphase als maßgeblich betrachtet und im Dossier dargestellt. Daten aus der offenen Behandlungsphase sind aufgrund des nach Entblindung erfolgten Therapiewechsels bei

---

<sup>17</sup> Ein finaler CSR lag zum Zeitpunkt der Dossiereinreichung noch nicht vor; zu den Ergebnissen der doppelblinden Behandlungsphase ist ein primärer CSR verfügbar, in dem die Ergebnisse der präspezifizierten primären Analyse berichtet werden.

Patient:innen aus dem Placebo-Arm nicht für die Beantwortung der Fragestellung der Nutzenbewertung geeignet.

## ***Ergebnisse***

### *Analysepopulationen*

Insgesamt wurden 96 Patient:innen in den Luspatercept-Arm und 49 Patient:innen in den Placebo-Arm randomisiert (ITT-Population). Alle randomisierten Patient:innen erhielten entsprechend der Randomisierung mindestens eine Dosis der Studientherapie. Die Safety-Population entspricht daher der ITT-Population.

### *Baseline-Charakteristika*

Die Mehrheit der randomisierten Patient:innen war weiblich (56,6%), weiß (60,0 %) und im Alter  $> 32 - \leq 50$  Jahre (46,2 %). Das mediane Alter betrug 40,0 Jahre. Der mediane BMI lag bei 21,6 kg/m<sup>2</sup>. Insgesamt wurden 62,1 % der Patient:innen in Nordamerika und Europa randomisiert, 11,7 % der Patient:innen im Mittleren Osten und die übrigen 26,2 % der Patient:innen in der Asien/Pazifik-Region.

Der Diagnose entsprechend wiesen die meisten Patient:innen eine reine Beta-Thalassämie auf (66,9 %). Bei den übrigen Patient:innen lag eine HbE/Beta-Thalassämie (26,9 %) oder eine Beta-Thalassämie kombiniert mit einer Alpha-Thalassämie (6,2 %) vor. Bei insgesamt 70,3 % der Patient:innen wurde einer der Genotypen  $\beta^0/\beta^0$ ,  $\beta^+/\beta^+$  oder  $\beta^+/\beta^0$ , jeweils ohne Alpha-Thalassämie, festgestellt.

Der Anteil der Patient:innen mit vorhergehender Splenektomie lag insgesamt bei 41,4 %.

Die Mehrheit der Patient:innen (57,9 %) wies einen Baseline-Hb-Wert  $< 8,5$  g/dl auf. Insgesamt zeigte sich eine für das Anwendungsgebiet typische niedrige Transfusionslast vor Studienbeginn. In den 24 Wochen vor Behandlungsbeginn benötigten 86,2 % der Patient:innen keine EK-Einheiten.

Insgesamt 69,0 % der Patient:innen wiesen einen ECOG-Status von 0 auf. Bei allen übrigen Patient:innen lag ein ECOG-Status von 1 vor.

Nur 35,9 % der Patient:innen wiesen zu Baseline normale LIC-Werte  $< 3$  mg/g d. w. auf. Insgesamt 30,3 % der Patient:innen erhielten in den 24 Wochen vor Studienbeginn eine Therapie mit Eisenchelatoren.

### *Behandlungs- und Beobachtungsdauer*

Bis zum primären Datenschnitt brachen 4,2 % der Patient:innen im Luspatercept-Arm und 2,0 % der Patient:innen im Placebo-Arm die Studie ab. Ein Therapieabbruch erfolgte bei 15,6 % der Patient:innen im Luspatercept-Arm und 63,3 % der Patient:innen im Placebo-Arm, wobei die Therapie von den meisten Patient:innen (92,7 % bzw. 71,4 %) bis mindestens Woche 48 fortgeführt wurde (siehe Tabelle 4-13).

Die mediane Beobachtungsdauer bis zum letzten Patientenkontakt war mit 106,3 Wochen bzw. 102,5 Wochen in beiden Studienarmen relativ ausgeglichen. Dagegen war die mediane Therapiedauer mit 99,7 Wochen im Luspatercept-Arm merklich länger als im Placebo-Arm mit 61,1 Wochen, wodurch sich auch die Beobachtungsdauer bei verschiedenen Endpunkten zwischen den Studienarmen unterscheidet (vgl. Abschnitt 4.3.1.3.1).

Tabelle 4-13: Studien- und Therapieabbrüche bis zum primären Datenschnitt

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)	Placebo + BSC <sup>(1)</sup> (N=49)	Insgesamt (N=145)
<b>Studienteilnahme</b>			
Studienabbruch [n (%)]	4 (4,2)	1 (2,0)	5 (3,4)
Beobachtungsdauer [Wochen] <sup>(2)</sup>			
Mittelwert (SD)	98,6 (26,10)	97,8 (24,63)	98,3 (25,54)
Median (Min; Max)	106,3 (19,9; 131,4)	102,5 (5,6; 130,9)	105,4 (5,6; 131,4)
<b>Studientherapie</b>			
Therapieabbruch [n (%)]	15 (15,6)	31 (63,3)	46 (31,7)
Therapie andauernd [n (%)]	81 (84,4)	18 (36,7)	99 (68,3)
≥ 24 Wochen abgeschlossen [n (%)]	92 (95,8)	44 (89,8)	136 (93,8)
≥ 48 Wochen abgeschlossen [n (%)]	89 (92,7)	35 (71,4)	124 (85,5)
Therapiedauer [Wochen] <sup>(3)</sup>			
Mittelwert (SD)	94,0 (29,76)	66,0 (35,02)	84,5 (34,20)
Median (Min; Max)	99,7 (15,0; 132,1)	61,1 (3,0; 121,9)	92,71 (3,0; 132,1)
Primärer Datenschnitt: 14.09.2020; ITT-Population bzw. Safety-Population			
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.			
<sup>(2)</sup> Zeit von der Randomisierung bis zum letzten Patientenkontakt.			
<sup>(3)</sup> Zeit von der ersten Dosis bis zur letzten Dosis + 20 Tage oder bis zum vorherigen Tod.			
BSC: <i>Best-Supportive-Care</i> ; Max: Maximum; Min: Minimum; SD: Standardabweichung			

### Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext

Informationen zur Charakterisierung der in Deutschland lebenden Patient:innen mit Beta-Thalassämie sind limitiert. Durch den Einschluss von Patient:innen in Griechenland, Italien, Libanon, Thailand, Vereinigtes Königreich und USA wurden vornehmlich weiße Patient:innen (60,0 %), jedoch auch ein bedeutender Anteil asiatischer Patient:innen (30,3 %) in die Studie eingeschlossen. In Deutschland haben Patient:innen mit Beta-Thalassämie meist ausländische Wurzeln: Neben Patient:innen, die überwiegend weißen Ethnien wie z. B. Italien oder Griechenland entstammen, haben viele Patient:innen eine asiatische Herkunft (Mittlerer Osten, Südostasien). Damit ist die Patientenpopulation in der Studie BEYOND eine adäquate Reflektion der betroffenen ethnischen Zugehörigkeiten in Deutschland [47-50]. Weiterhin ist mit vielen Studienzentren in westlichen Ländern ein mit Deutschland vergleichbarer Standard der Gesundheitsversorgung sichergestellt.

Aufgrund der genetischen Ursache ist davon auszugehen, dass weibliche und männliche Patient:innen gleichermaßen von Beta-Thalassämie betroffen sind [47, 48]. Entsprechend waren in der Studie Frauen und Männer zu etwa gleichen Teilen eingeschlossen (63 Männer (43,4 %), 82 Frauen (56,6 %)). Die tendenziell etwas höhere Anzahl an Frauen spiegelt sich auch in der im Rahmen der Dossiererstellung in Auftrag gegebenen Auswertungen von Routinedaten der gesetzlichen Krankenkassen (41,07 % Männer, 58,93 % Frauen; siehe Modul 3 C) wider [51].

Aufgrund der vorliegenden Patientencharakteristika und des multizentrischen Studiendesigns wird von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ausgegangen.

#### 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 4-14: 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			
BEYOND	ja	ja	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Die Studie BEYOND wurde randomisiert und doppelblind durchgeführt. Um die adäquate Gruppenzuteilung und Wahrung der Verblindung der Studienteilnehmer:innen, behandelnden Personen sowie des pharmazeutischen Unternehmers zu gewähren, erfolgten die Randomisierung, Arzneimittelausgabe, Dosistitration/-reduktion und Therapieabbrüche unterstützt durch ein *Interactive Response Technology*-(IRT-)System. Ausgenommen von der Verblindung waren lediglich ausgewählte Einzelpersonen (z. B. Apotheker:innen des Studienzentrums), die eine adäquate Herstellung der Luspatercept- bzw. Placebo-Gaben sicherstellten.

Die Erhebung von präspezifizierten Endpunkten sowie die Darstellung der Ergebnisse erfolgte wie vorab definiert. Der Patientenfluss ist nachvollziehbar dargestellt und es wurden keine außergewöhnlichen, das Verzerrungspotenzial der Studie maßgeblich beeinflussenden Unregelmäßigkeiten berichtet. Dementsprechend liegen weder Hinweise auf eine ergebnisabhängige Berichterstattung noch sonstige Aspekte vor, die die Ergebnissicherheit der Studie in Frage stellen würden.

Angaben zum Verzerrungspotenzial auf Endpunktebene finden sich in Anhang 4-F.

Insgesamt ist das Verzerrungspotenzial der Studie BEYOND niedrig und die Studie somit geeignet, um Ergebnisse mit hoher Aussagesicherheit zu liefern.

#### 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 4-15: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

Studie	Mortalität	Morbidität					Lebensqualität		Sicherheit <sup>(1)</sup>
		Vermeidung von Transfusionen	Kontrolle der Eisenüberladung (ergänzend dargestellt)	Symptomatik gemäß NTDT-PRO	Symptomatik gemäß PGI-S und PGI-C	Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest	SF-36	FACIT-F	
BEYOND	ja	ja	ja	ja	ja	ja	ja	ja	ja

<sup>(1)</sup> Dargestellt anhand von UE, einschließlich schwerer UE (CTCAE Grad  $\geq 3$ ), schwerwiegender UE und zum Therapieabbruch führender Ereignisse.  
CTCAE: Common Terminology Criteria for Adverse Events; FACIT-F: Functional Assessment of Chronic Illness Therapy–Fatigue; NTDT-PRO: Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes; PGI-C: Patient Global Impression of Change; PGI-S: Patient Global Impression of Severity; SF-36: Short Form (36) Health Survey; UE: unerwünschtes Ereignis

#### 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  $\geq 3$ ) 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.

#### 4.3.1.3.1.1 Mortalität

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

Tabelle 4-16: Operationalisierung des Endpunkts „Mortalität“

Studie	Operationalisierung
BEYOND	<p>Die Mortalität wurde in der Studie BEYOND anhand von Todesfällen jeglicher Ursache erhoben. Es werden somit alle während der Studie verstorbenen Patient:innen gezählt.</p> <p>Als Behandlungseffekt wird das unter Berücksichtigung der für die Randomisierung verwendeten Stratifizierungsfaktoren mittels Mantel-Haenszel-Methode geschätzte relative Risiko betrachtet. Die Analyse basiert auf der ITT-Population.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> Anteil der Todesfälle in jedem Studienarm</li> </ul>
ITT: Intention-to-Treat	

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 4-17: Bewertung des Verzerrungspotenzials für den Endpunkt „Mortalität“ in der Studie BEYOND

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

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt „Mortalität“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung oder sonstige Aspekte, die

zu einer Verzerrung führen könnten, liegen nicht vor. Der Eintritt des Todes jeglicher Ursache ist ein objektives Kriterium, dessen Feststellung frei von Subjektivität erfolgt.

Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

*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.*

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

Tabelle 4-18: Ergebnisse für den Endpunkt „Mortalität“ – Anteil der Todesfälle aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N	Patient:innen mit Ereignis n (%)	N	Patient:innen mit Ereignis n (%)	RD <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	OR <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	RR <sup>(2)</sup> [95 %-KI]; p-Wert <sup>(3)</sup>
<b>Anteil der verstorbenen Patient:innen</b>							
Todesfälle	96	0	49	0	N.A.	N.A.	N.A.
Primärer Datenschnitt: 14.09.2020; ITT-Population <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche. <sup>(3)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet. BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; OR: Odds Ratio; RD: Risiko Differenz; RR: Relatives Risiko							

In der Studie BEYOND verstarben bis zum maßgeblichen Datenschnitt auf Basis der doppelblinden Behandlungsphase keine Patient:innen.

*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.*

### **Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

### **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt Mortalität liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### **4.3.1.3.1.2 Morbidität – Vermeidung von Transfusionen**

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

Tabelle 4-19: Operationalisierung des Endpunkts „Morbidität – Vermeidung von Transfusionen“

<b>Studie</b>	<b>Operationalisierung</b>
BEYOND	<p>Als Hauptanalyse wird die Zeit bis zur ersten Transfusion betrachtet. Ergänzend dargestellt werden Auswertungen zur Transfusionsfreiheit, zur Anzahl benötigter Transfusionen, zur mittleren Änderung der Hb-Werte, zum mittleren Verlauf der Hb-Werte, zum erythroiden Ansprechen sowie zur Dauer des erythroiden Ansprechens.</p> <p>Transfusionen wurden in der Studie BEYOND während des Screenings, d. h. in den 4 Wochen vor der ersten Dosis, sowie für mindestens 24 Wochen retrospektiv vor der ersten Dosis und ab Therapiebeginn bis 9 Wochen nach der letzten Dosis erhoben.</p> <p>Die Zeit bis zur ersten Transfusion ist definiert als die Zeit von der ersten Dosis der Studienmedikation bis zur ersten EK-Transfusion während der Behandlungsphase. Die Auswertung erfolgte anhand einer Ereigniszeitanalyse. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne eine EK-Transfusion während der Behandlungsphase wurden zum Zeitpunkt des betrachteten Datenschnitts bzw. 20 Tage nach der letzten Dosis oder bei vorherigem Studienabbruch/Tod, je nachdem welches Ereignis früher eintrat, zensiert.</p> <p>Eine Transfusionsfreiheit von <math>\geq 24</math> Wochen liegt bei einer Patientin oder einem Patienten vor, wenn diese bzw. dieser während der betrachteten Behandlungsphase in einem kontinuierlichen Intervall von <math>\geq 24</math> Wochen keine EK-Transfusionen erhalten hat. Die Auswertung basiert auf dem Anteil der Patient:innen mit einer Transfusionsfreiheit über <math>\geq 24</math> Wochen während der Behandlungsphase im Zeitraum (i) bis Woche 24, (ii) bis Woche 48 und (iii) bis zum primären Datenschnitt. Analog hierzu erfolgte darüber hinaus eine Auswertung der Transfusionsfreiheit</p>

Studie	Operationalisierung
	<p>≥ 48 Wochen während der Behandlungsphase im Zeitraum (i) bis Woche 48 und (ii) bis zum primären Datenschnitt. Als Schätzer des Behandlungseffekts wird jeweils das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.</p> <p>Die Anzahl benötigter Transfusionen (0, 1, 2, usw.) wird deskriptiv als kategorielle Variable anhand des Patientenanteils pro Kategorie dargestellt.</p> <p>Die Bestimmung der Hb-Werte erfolgte durch ein Zentrallabor für zwei Zeitpunkte (im Abstand von ≥ 1 Woche) während des Screenings, d. h. in den 4 Wochen vor der ersten Dosis, sowie für jede Behandlungsvisite (gemessen vor der Dosis) und zusätzlich für Woche 13, 19, 24, 37, 42 und 48 (sofern die Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der Behandlungsvisiten stattfand), sowie für eine Messung 9 Wochen nach der letzten Dosis.</p> <p>Hb-Messungen innerhalb von 21 Tagen nach einer EK-Transfusion können durch die Transfusion verzerrt sein und wurden deshalb von den Auswertungen ausgeschlossen.</p> <p>Die Änderung der Hb-Werte während der Behandlungsphase im Zeitraum bis Woche 48 wurde für jeden Patienten bzw. jede Patientin auf Basis des Mittelwerts aller Hb-Messungen der Behandlungsphase bis Woche 48 im Vergleich zum Baseline-Hb-Wert (= Mittelwert der Hb-Messungen in den 4 Wochen vor der ersten Dosis) ermittelt. Die Auswertung erfolgte mittels ANCOVA unter Berücksichtigung des Baseline-Hb-Werts (als stetige Variable), des Behandlungsarms und des Baseline-NTDT-PRO-Gesamtscores in der Domäne Müdigkeit/Schwäche (≥ 3 Punkte vs. &lt; 3 Punkte) als unabhängige Variablen. Analog hierzu wurde auch die mittlere Änderung der Hb-Werte während der Behandlungsphase im Zeitraum bis Woche 24 ausgewertet.</p> <p>Zur graphischen Darstellung des Behandlungsverlaufs wurden die Hb-Werte der Patient:innen anhand des arithmetischen Mittelwerts pro Behandlungsgruppe jeweils in aufeinanderfolgenden 3-Wochen-Intervallen ab Woche 2 (Woche 2-4, Woche 5-7, Woche 8-10, usw.) zusammengefasst. Bei Patient:innen mit mehr als einer Messung innerhalb eines Intervalls wurde der Durchschnitt dieser Messungen herangezogen.</p> <p>Als erythroides Ansprechen wurde ein Anstieg von ≥ 1,0 g/dl im Mittelwert der Hb-Messungen über 12 Wochen im Vergleich zum Baseline-Hb-Wert (= Mittelwert der Hb-Messungen in den 4 Wochen vor der ersten Dosis) definiert. Für längere Zeiträume wurde der Hb-Anstieg in Form eines gleitenden Mittelwerts über 12 Wochen (= 84 Tage) betrachtet. D. h. ein erythroides Ansprechen über einen bestimmten Zeitraum (z. B. 24 Wochen) liegt vor, wenn der Mittelwert der Hb-Messungen einer Patientin oder eines Patienten in jedem möglichen 12-Wochen-Intervall innerhalb dieses Zeitraums (also von Tag x bis x+84, von Tag x+1 bis x+85, usw.) im Vergleich zum Baseline-Wert um ≥ 1,0 g/dl erhöht war; typischerweise umfassen 12-Wochen-Intervalle 4 Hb-Messungen. Dargestellt wird der Anteil der Patient:innen mit einem erythroiden Ansprechen über mindestens 24 Wochen (und ergänzend mindestens 48 Wochen). Die Auswertung erfolgte auf Basis der Behandlungsphase im Zeitraum (i) bis Woche 24, (ii) bis Woche 48 und (iii) bis zum primären Datenschnitt. Als Schätzer des Behandlungseffekts wird jeweils das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.</p> <p>Die Dauer des erythroiden Ansprechens wurde in Form der längsten Dauer anhand einer Ereigniszeitanalyse ausgewertet. Die längste Dauer des erythroiden Ansprechens entspricht dem längsten Zeitraum, in dem die Hb-Werte einer Patientin bzw. eines Patienten im gleitenden Mittelwert über 12 Wochen durchgängig um ≥ 1,0 g/dl erhöht waren. Diese Analyse umfasst nur Patient:innen, die mindestens einmal ein erythroides Ansprechen erreichten. Ein Ereignis liegt bei Patient:innen vor, die ihr längstes erythroides Ansprechen zum Datenschnitt bzw. bis 20 Tage nach Therapieabbruch oder Studienabbruch/Tod (je nachdem, was früher eintrat) wieder verloren hatten. Patient:innen, deren erythroides Ansprechen zu diesem Zeitpunkt noch andauerte, wurden zensuriert.</p> <p>Alle Analysen zu diesem Endpunkt basieren auf der ITT-Population.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> Zeit bis zur ersten Transfusion</li> </ul>

Studie	Operationalisierung
	<ul style="list-style-type: none"> <li>• <b>Zusatzanalysen:</b> <ul style="list-style-type: none"> <li>○ Anteil der Patient:innen mit einer Transfusionsfreiheit über mindestens 24 bzw. 48 Wochen</li> <li>○ Anzahl benötigter Transfusionen</li> <li>○ Mittlere Änderung der Hb-Werte (ANCOVA)</li> <li>○ Line-Plot zum mittleren Verlauf der Hb-Werte</li> <li>○ Erythroides Ansprechen über mindestens 24 bzw. 48 Wochen</li> <li>○ Längste Dauer des erythroiden Ansprechens</li> </ul> </li> </ul>
ANCOVA: <i>Analysis of covariance</i> ; BSC: <i>Best-Supportive-Care</i> ; EK: Erythrozytenkonzentrat; Hb: Hämoglobin; HR: Hazard Ratio; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i>	

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

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

Tabelle 4-20: Bewertung des Verzerrungspotenzials für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ in der Studie BEYOND

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

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt „Vermeidung von Transfusionen“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch eine Ereigniszeitanalyse adressiert (siehe Tabelle 4-21). Die Auswertung umfasst den Behandlungszeitraum bis einschließlich 20 Tage nach der letzten Dosis. In Bezug auf Patient:innen, die die Therapie abbrachen, ist bei der betrachteten Ereigniszeitanalyse grundsätzlich die Möglichkeit potenziell informativer Zensurierungen nicht auszuschließen. In Anbetracht der Größe des beobachtenden Effekts der Ereigniszeitanalyse sowie der konsistenten Responderanalysen ist jedoch davon auszugehen, dass dies im Ergebnis keine entscheidende Rolle spielt.

Tabelle 4-21: Beobachtungszeiten für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ in der Studie BEYOND

<b>Studie BEYOND Beobachtungszeit in Wochen</b>	<b>Luspatercept + BSC <sup>(1)</sup> (N=96)</b>	<b>Placebo + BSC <sup>(1)</sup> (N=49)</b>
Mittelwert (SD)	94,0 (29,8)	66,0 (35,0)
Median (Min; Max)	99,7 (15; 132)	61,1 (3; 122)
Primärer Datenschnitt: 14.09.2020; ITT-Population		
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.		
BSC: <i>Best-Supportive-Care</i> ; ITT: Intention-to-Treat; Max: Maximum; Min: Minimum; SD: Standardabweichung		

Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

*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.*

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

**Hauptanalyse: Zeit bis zur ersten Transfusion**

Tabelle 4-22: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Zeit bis zur ersten Transfusion aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Wochen) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Wochen) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
<b>Zeit bis zur ersten Transfusion</b>								
Zeit bis zur ersten Transfusion	96	14 (14,6)	N.A. [N.A.; N.A.]	49	21 (42,9)	86,14 [43,86; N.A.]	0,23 [0,11; 0,46]	< 0,0001
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.								
<sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten Transfusion.								
<sup>(3)</sup> Cox-Regressionsmodell stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(4)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

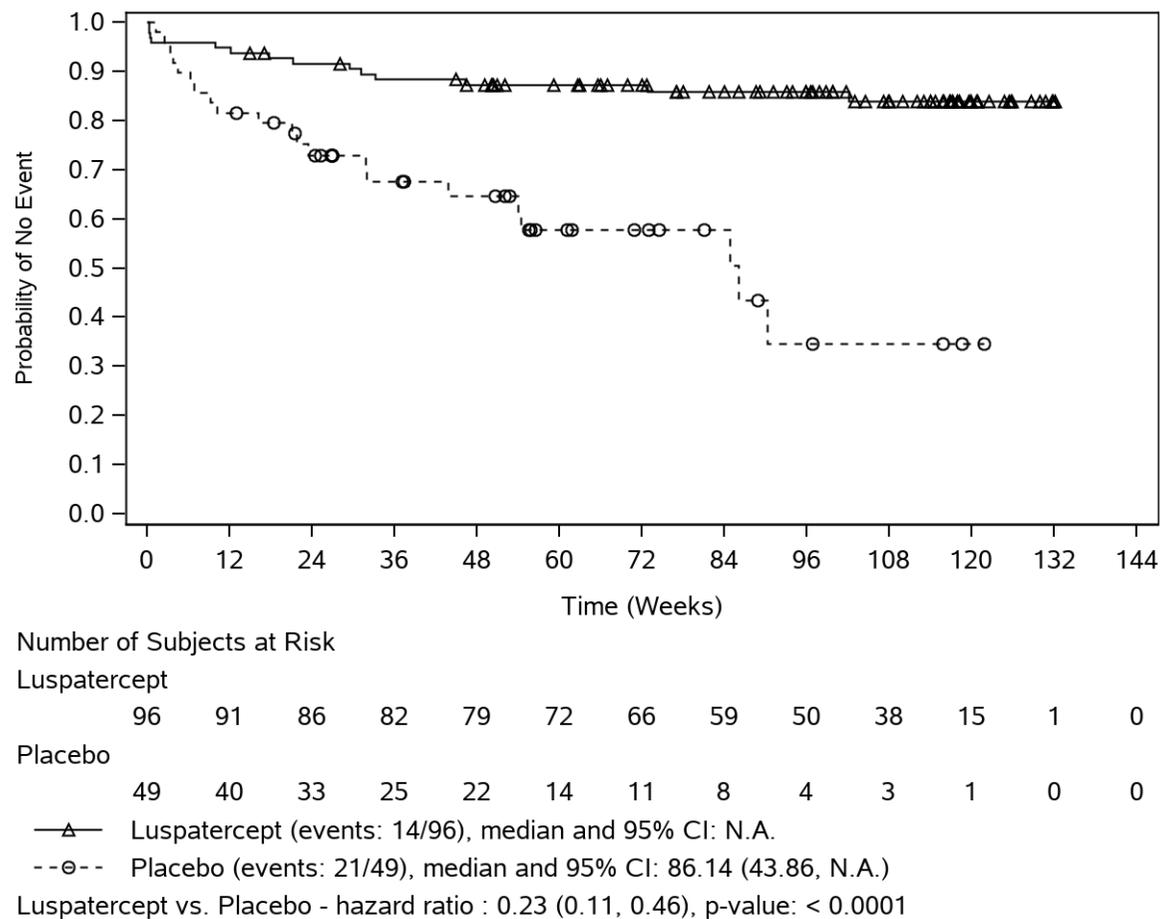


Abbildung 4-3: Kaplan-Meier Plot der Zeit bis zur ersten Transfusion

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

Für den Endpunkt „Vermeidung von Transfusionen“ zeigte sich in der Studie BEYOND ein statistisch signifikanter Effekt zugunsten von Luspatercept.

In der Ereigniszeitanalyse ergab sich für den Vergleich von Luspatercept gegenüber Placebo ein HR von 0,23 (95 %-KI [0,11; 0,46];  $p < 0,0001$ ). Dies entspricht einer signifikanten Reduktion des Risikos für eine Transfusion um 77 %. Diese langfristige Vermeidung von Transfusionen ging gleichzeitig mit einer Steigerung und Stabilisierung der Hb-Werte und somit einer anhaltenden Linderung der Anämie einher (siehe nachfolgende Zusatzanalysen zur Änderung der Hb-Werte und zum erythroiden Ansprechen). Wie in Abbildung 4-3 ersichtlich, wurde der Median der Zeit bis zur ersten Transfusion im Luspatercept-Arm zum Zeitpunkt des Datenschnitts noch nicht erreicht. Bei Patient:innen im Placebo-Arm betrug die mediane Zeit bis zur ersten Transfusion 86,14 Wochen (95 %-KI [43,86; N.A.]).

Weitere Zusatzanalysen sind im Folgenden dargestellt.

## Zusatzanalysen

### *Transfusionsfreiheit über mindestens 24 bzw. 48 Wochen*

Tabelle 4-23: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Transfusionsfreiheit über mindestens 24 bzw. 48 Wochen aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N	Patient:innen mit Ansprechen n (%)	N	Patient:innen mit Ansprechen n (%)	RD <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	OR <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	RR <sup>(2)</sup> [95 %-KI]; p-Wert <sup>(3)</sup>
<b>Transfusionsfreiheit ≥ 24 Wochen</b>							
Bis Woche 24	96	86 (89,6)	49	33 (67,3)	22,24 [4,98; 38,62]	4,32 [1,72; 10,81]	1,32 [1,08; 1,61]; <b>0,0071</b>
Bis Woche 48	96	90 (93,8)	49	37 (75,5)	18,24 [0,80; 34,69]	4,8 [1,67; 13,80]	1,24 [1,05; 1,47]; <b>0,0121</b>
Bis Datenschnitt	96	91 (94,8)	49	38 (77,6)	17,24 [-0,23; 33,70]	5,26 [1,70; 16,23]	1,22 [1,04; 1,43]; <b>0,0129</b>

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N	Patient:innen mit Ansprechen n (%)	N	Patient:innen mit Ansprechen n (%)	RD <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	OR <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	RR <sup>(2)</sup> [95 %-KI]; p-Wert <sup>(3)</sup>
<b>Transfusionsfreiheit ≥ 48 Wochen</b>							
Bis Woche 48	96	79 (82,3)	49	22 (44,9)	37,39 [20,93; 52,96]	5,96 [2,69; 13,17]	1,81 [1,32; 2,49]; <b>0,0003</b>
Bis Datenschnitt	96	83 (86,5)	49	24 (49,0)	37,48 [20,93; 53,06]	6,68 [2,96; 15,05]	1,76 [1,31; 2,37]; <b>0,0002</b>
ITT-Population; Primärer Datenschnitt vom 14.09.2020							
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.							
<sup>(2)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche							
<sup>(3)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet.							
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; OR: Odds Ratio; RD: Risikodifferenz; RR: Relatives Risiko							

Eine Transfusionsfreiheit über 24 Wochen erreichten während der Behandlungsphase insgesamt 94,8 % der Patient:innen im Luspatercept-Arm und 77,6 % der Patient:innen im Placebo-Arm. Eine Transfusionsfreiheit über 48 Wochen erreichten insgesamt 86,5 % der Patient:innen im Luspatercept-Arm und 49,0 % der Patient:innen im Placebo-Arm. Damit zeigt sich zwischen den Behandlungsarmen jeweils ein deutlicher, statistisch signifikanter Unterschied zugunsten von Luspatercept (p = 0,0129 bzw. p = 0,0002; siehe Tabelle 4-23). Diese langfristige Transfusionsfreiheit ging gleichzeitig mit einer Steigerung und Stabilisierung der Hb-Werte und somit einer anhaltenden Linderung der Anämie einher (siehe nachfolgende Zusatzanalysen zur Änderung der Hb-Werte und zum erythroiden Ansprechen).

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

### Anzahl benötigter Transfusionen

Tabelle 4-24: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Anzahl benötigter Transfusionen aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)	Placebo + BSC <sup>(1)</sup> (N=49)
	n (%)	n (%)
<b>Anzahl benötigter Transfusionen bis Woche 24</b>		
0	88 (91,7)	36 (73,5)
1	4 (4,2)	8 (16,3)
2	0	4 (8,2)
3	3 (3,1)	0
4	1 (1,0)	0
5	0	1 (2,0)
<b>Anzahl benötigter Transfusionen bis Woche 48</b>		
0	84 (87,5)	33 (67,3)
1	7 (7,3)	5 (10,2)
2	0	7 (14,3)
3	1 (1,0)	2 (4,1)
4	3 (3,1)	1 (2,0)
6	1 (1,0)	0
9	0	1 (2,0)

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)	Placebo + BSC <sup>(1)</sup> (N=49)
	n (%)	n (%)
<b>Anzahl benötigter Transfusionen bis Datenschnitt</b>		
0	82 (85,4)	28 (57,1)
1	4 (4,2)	6 (12,2)
2	2 (2,1)	6 (12,2)
3	3 (3,1)	4 (8,2)
4	4 (4,2)	1 (2,0)
6	0	1 (2,0)
7	0	1 (2,0)
8	0	1 (2,0)
12	1 (1,0)	0
19	0	1 (2,0)
ITT-Population; Primärer Datenschnitt vom 14.09.2020 <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. BSC: <i>Best-Supportive-Care</i> ; ITT: Intention-to-Treat		

Die Auswertung zur Anzahl benötigter Transfusionen innerhalb der Behandlungsphase bestätigt den geringeren Bedarf an Transfusionen im Luspatercept-Arm. Weiterhin wird ersichtlich, dass Patient:innen im Placebo-Arm deutlich häufiger mehr als eine Transfusion benötigten. Während im Luspatercept-Arm insgesamt nur 10,4 % der Patient:innen mehr als eine Transfusion erhielten, waren es im Placebo-Arm 30,7 % der Patient:innen (siehe Tabelle 4-24).

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

**Mittlere Änderung der Hb-Werte (ANCOVA)**

Tabelle 4-25: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – Änderung der Hb-Werte gegenüber Baseline aus der Studie BEYOND (Zusatzanalyse)

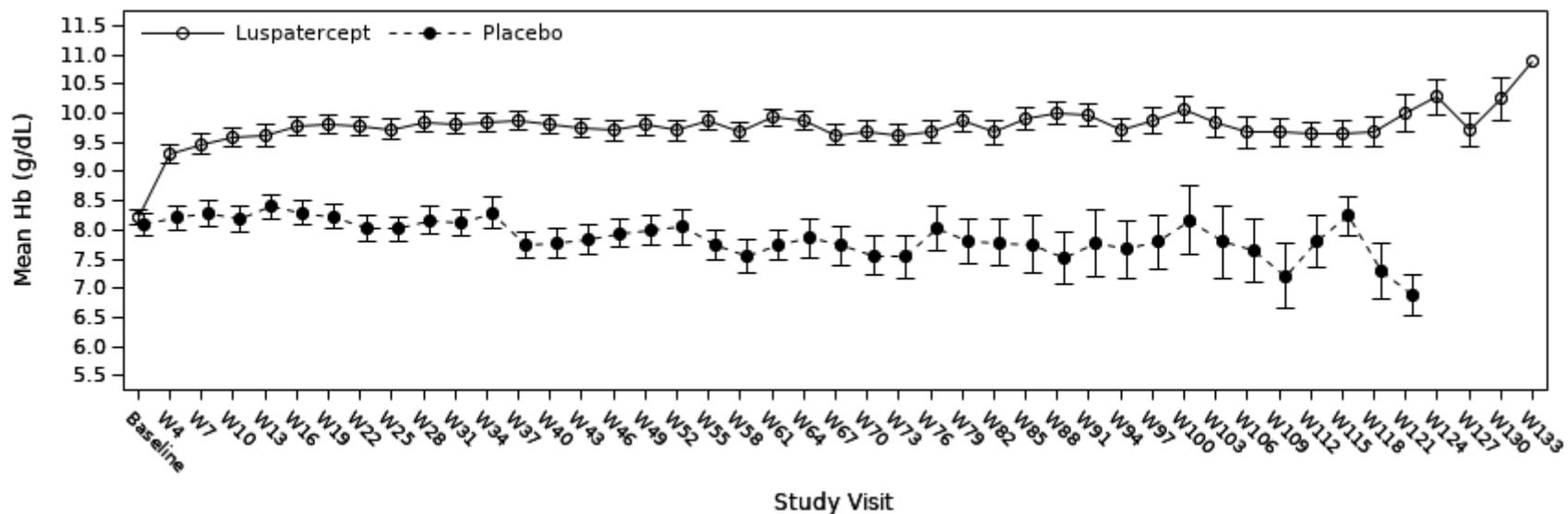
Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>	
	N	Baseline Mittel (SD) <sup>(2)</sup>	Änderung gegenüber Baseline Mittel (SE) <sup>(3)</sup>	N	Baseline Mittel (SD) <sup>(2)</sup>	Änderung gegenüber Baseline Mittel (SE) <sup>(3)</sup>	Mittlere Differenz [95 %-KI]; p-Wert <sup>(3)</sup>	Hedges' g [95 %-KI] <sup>(4)</sup>
<b>Änderung des Hb gegenüber Baseline (g/dl)</b>								
Mittlere Änderung bis Woche 24 (g/dl)	96	8,2 (1,22)	1,40 (0,071)	49	8,09 (1,30)	0,07 (0,096)	1,33 [1,10; 1,55]; < 0,0001	2,04 [1,62; 2,46]
Mittlere Änderung bis Woche 48 (g/dl)	96	8,2 (1,22)	1,44 (0,072)	49	8,09 (1,30)	0,10 (0,098)	1,35 [1,12; 1,57]; < 0,0001	2,04 [1,62; 2,46]
ITT-Population; Primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.								
<sup>(2)</sup> Arithmetischer Mittelwert basierend auf den durchschnittlichen Hb-Werten in den 4 Wochen vor der ersten Dosis								
<sup>(3)</sup> ANCOVA basierend auf der Änderung gegenüber Baseline, adjustiert für den Baseline-Hb (als stetiger Wert) und den Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(4)</sup> Hedges' g abgeleitet aus dem Modellschätzer der mittleren Differenz								
ANCOVA: <i>analysis of covariance</i> ; BSC: <i>Best-Supportive-Care</i> ; Hb: Hämoglobin; ITT: Intention-to-Treat; KI: Konfidenzintervall; SD: Standardabweichung; SE: Standardfehler								

Im Durchschnitt ergab sich über die Behandlungsphase bis Woche 48 im Luspatercept-Arm ein mittlerer Anstieg der Hb-Werte um 1,44 g/dl während es im Placebo-Arm nur 0,10 g/dl waren. Die Differenz zwischen den beiden Armen ist statistisch signifikant (mittlere Differenz: 1,35 [1,12; 1,57]; p < 0,0001; siehe Tabelle 4-25).

Die Erhöhung des mittleren Hb-Spiegels im Luspatercept-Arm blieb über den gesamten Behandlungsverlauf bis zum Datenschnitt stabil (siehe Abbildung 4-4). Diese Ergebnisse verdeutlichen, dass Luspatercept eine Verbesserung der Erythropoese bewirkt, wodurch

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

Patient:innen im Luspatercept-Arm gleichzeitig Transfusionen vermeiden konnten und langfristig von einer Linderung ihrer Anämie profitierten.



Primärer Datenschnitt vom 14.09.2020; ITT-Population

Abbildung 4-4: Line-Plot zum mittleren Verlauf der Hb-Werte

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

### ***Erythroides Ansprechen über mindestens 24 bzw. 48 Wochen***

Tabelle 4-26: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – erythroides Ansprechen über mindestens 24 bzw. 48 Wochen aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N	Patient:innen mit Ansprechen n (%)	N	Patient:innen mit Ansprechen n (%)	RD <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	OR <sup>(2)</sup> [95 %-KI] <sup>(3)</sup>	RR <sup>(2)</sup> [95 %-KI]; p-Wert <sup>(3)</sup>
<b>Erythroides Ansprechen ≥ 24 Wochen<sup>(4)</sup></b>							
Bis Woche 24	96	63 (65,6)	49	0	65,63 [50,83; 78,33]	N.A.	N.A.
Bis Woche 48	96	76 (79,2)	49	1 (2,0)	77,13 [63,38; 86,97]	133,92 [19,06; 940,99]	37,25 [5,53; 250,72]; <b>0,0002</b>
Bis Datenschnitt	96	79 (82,3)	49	2 (4,1)	78,21 [64,83; 88,23]	87,96 [20,64; 374,85]	20,13 [5,15; 78,71]; <b>&lt; 0,0001</b>
<b>Erythroides Ansprechen ≥ 48 Wochen<sup>(4)</sup></b>							
Bis Woche 48	96	55 (57,3)	49	0	57,29 [41,78; 71,18]	N.A.	N.A.
Bis Datenschnitt	96	67 (69,8)	49	0	69,79 [55,43; 81,75]	N.A.	N.A.
ITT-Population; Primärer Datenschnitt vom 14.09.2020							
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.							
<sup>(2)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche							
<sup>(3)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet.							
<sup>(4)</sup> Anstieg des mittleren Hb-Werts der Patientin bzw. des Patienten um ≥ 1,0 g/dl über mindestens 24 bzw. 48 Wochen während der Behandlungsphase (durchgängig beurteilt am gleitenden Mittelwert über 12 Wochen) im Vergleich zum mittleren Hb-Wert der 4 Wochen vor der ersten Dosis							
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; OR: Odds Ratio; RD: Risikodifferenz; RR: Relatives Risiko							

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

Ein erythroides Ansprechen über mindestens 24 bzw. 48 Wochen während der Behandlungsphase wurde fast ausschließlich von Patient:innen im Luspatercept-Arm erreicht. So lag ein erythroides Ansprechen über 24 Wochen bis zum Datenschnitt bei 82,3 % der Patient:innen im Luspatercept-Arm und 4,1 % der Patient:innen im Placebo-Arm vor. Damit zeigte sich zwischen den Behandlungsarmen ein deutlicher, statistisch signifikanter Unterschied zugunsten von Luspatercept ( $p < 0,0001$ ; siehe Tabelle 4-26). Dieser Vorteil wurde auch in Bezug auf ein erythroides Ansprechen über mindestens 48 Wochen beibehalten. Ein entsprechend langes erythroides Ansprechen zeigten bis zum Datenschnitt 69,8 % der Patient:innen im Luspatercept-Arm, während keine Patient:innen im Placebo-Arm eine solche Verbesserung erreichten.

### ***Längste Dauer des erythroiden Ansprechens***

Tabelle 4-27: Ergebnisse für den Endpunkt „Morbidity – Vermeidung von Transfusionen“ – längste Dauer des erythroiden Ansprechens aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis <sup>(3)</sup> n (%)	Median in Wochen <sup>(4)</sup> [95 %-KI]	N <sup>(2)</sup>	Patient:innen mit Ereignis <sup>(3)</sup> n (%)	Median in Wochen <sup>(4)</sup> [95 %-KI]	Hazard Ratio <sup>(5)</sup> [95 %-KI]	p-Wert <sup>(6)</sup>
<b>Längste Dauer des erythroiden Ansprechens</b>								
Längstes erythroides Ansprechen	88	39 (44,3)	102,86 [80,43; N.A.]	11	8 (72,7)	20,86 [17,29; 29,86]	0,10 [0,04; 0,26]	< 0,0001
ITT-Population; Primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Die Analyse umfasst nur Patient:innen, die mindestens einmal ein erythroides Ansprechen erreichten, d.h. Patient:innen mit einem Anstieg des mittleren Hb-Werts während der Behandlungsphase (gleitender Mittelwert über 12-Wochen) um $\geq 1,0$ g/dl im Vergleich zum mittleren Hb-Wert der 4 Wochen vor der ersten Dosis. <sup>(3)</sup> Ein Ereignis liegt bei Patient:innen vor, die ihr längstes erythroides Ansprechen zum Datenschnitt bzw. bis 20 Tage nach Therapieabbruch oder Studienabbruch/Tod (je nachdem, was früher eintrat) wieder verloren hatten. Patient:innen, deren erythroides Ansprechen zu diesem Zeitpunkt noch andauerte, wurden zensiert. <sup>(4)</sup> Mediane Zeit nach Kaplan-Meier <sup>(5)</sup> Cox-Regressionsmodell stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche <sup>(6)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche BSC: <i>Best-Supportive-Care</i> ; ITT: Intention-to-Treat; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

---

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

Bei den Patient:innen, die bezogen auf ein 12-Wochen-Intervall mindestens einmal während der Behandlungsphase bis zum Datenschnitt ein erythroides Ansprechen erreichten, lag die längste Dauer des erythroiden Ansprechens im Luspatercept-Arm im Median bei 102,86 Wochen gegenüber 20,86 Wochen im Placebo-Arm (siehe Tabelle 4-27).

### **Zusammenfassung der Ergebnisse zum Endpunkt Vermeidung von Transfusionen**

Die Analysen zum Endpunkt „Vermeidung von Transfusionen“ zeigen durchgängig statistisch signifikante Vorteile zugunsten von Luspatercept, sowohl in der Hauptanalyse (Zeit bis zur ersten Transfusion) als auch im Rahmen der betrachteten Zusatzanalysen. Neben der Vermeidung von Transfusionen führte die Therapie mit Luspatercept gegenüber Placebo zu einer signifikanten Steigerung sowie einer langfristigen Stabilisierung der Hb-Werte und somit zu einer anhaltenden Linderung der Anämie.

Dies bedeutet erstens, dass die Patient:innen langfristig von eingesparten EK-Transfusionen aufgrund der Vermeidung von Transfusionsrisiken, Folgekomplikationen und einer weiteren Eisenüberladung profitieren können. Zweitens ist vor dem chronischen Hintergrund der NTDT der Erhalt der Transfusionsunabhängigkeit sowie die Verhinderung eines Übergangs in eine transfusionsabhängige Beta-Thalassämie, die mit regelmäßiger Transfusionstherapie und einem erhöhten Behandlungsaufwand, insbesondere einem zeitlichen Aufwand einhergeht, von großer Bedeutung für die Betroffenen. Drittens führt die durch Luspatercept bedingte Steigerung und Stabilisierung der Hb-Werte (Verbesserung der Erythropoese), nicht nur zu einer langanhaltenden Vermeidung von Transfusionen, sondern ermöglicht im Weiteren auch eine Reduktion Anämie-assoziiierter Symptomatik und folglich auch eine spürbare Verbesserung im Befinden der Patient:innen.

*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.*

### **Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

### **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Morbidität - Vermeidung von Transfusionen“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### 4.3.1.3.1.3 Morbidität – Kontrolle der Eisenüberladung

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

Tabelle 4-28: Operationalisierung des Endpunkts „Morbidität – Kontrolle der Eisenüberladung“

Studie	Operationalisierung
BEYOND	<p>Die Kontrolle der Eisenüberladung wird anhand der Änderung des sFerritin-Spiegels dargestellt. Die Messung der sFerritin-Werte erfolgte in der Studie BEYOND zu einem Zeitpunkt während des Screenings, d. h. in den 4 Wochen vor Behandlungsbeginn, sowie für mindestens 24 Wochen retrospektiv vor Behandlungsbeginn, zum Zeitpunkt jeder Behandlungsvisite in der doppelblinden Behandlungsphase und zusätzlich zu Woche 24 und 48 (sofern die Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der Behandlungsvisiten stattfand), sowie 9 Wochen nach der letzten Dosis.</p> <p>Dargestellt wird die Änderung des mittleren sFerritin-Spiegels in den letzten 24 Wochen der Behandlungsphase bis Woche 48 gegenüber dem Baseline-Wert (= Mittelwert der sFerritin-Messungen in den 24 Wochen vor der ersten Dosis).</p> <p>Die Auswertung erfolgte mittels ANCOVA, wobei eine Adjustierung für den Baseline-Wert sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren vorgenommen wurde. Zur Quantifizierung des Gruppenunterschieds wird die adjustierte Mittelwertdifferenz sowie der entsprechende Schätzer von Hedges' g betrachtet.</p> <p>Ergänzend zum sFerritin wurden Auswertungen zur mittleren Änderung der LIC und des Eisenchelator-Bedarfs durchgeführt.</p> <p>Die Messung der LIC (mittels MRT) erfolgte in der Studie BEYOND zu einem Zeitpunkt während des Screenings, d. h. in den 4 Wochen vor der ersten Dosis (sofern innerhalb der 12 Wochen vor Unterzeichnung der Einverständniserklärung kein durchgeführtes MRT zur Verfügung stand), sowie zu Woche 24 und 48. Angaben zu Eisenchelatoren wurden während des Screenings sowie für 24 Wochen retrospektiv vor der ersten Dosis und ab Therapiebeginn bis 9 Wochen nach der letzten Dosis erhoben.</p> <p>Die mittlere Änderung der LIC zu Woche 48 (und ergänzend Woche 24) gegenüber der letzten Messung vor Behandlungsbeginn (= Baseline) wurde mittels ANCOVA ausgewertet. Die Änderung des Eisenchelator-Bedarfs wurde pro Wirkstoff anhand der Änderung der durchschnittlichen täglichen Dosis in den letzten 24 Wochen der Behandlungsphase bis Woche 48 gegenüber den 24 Wochen vor Behandlungsbeginn (= Baseline) mittels ANCOVA ausgewertet. Hierbei wurden alle Patient:innen berücksichtigt, die eine Eisenchelator-Monotherapie erhielten und ihren Wirkstoff im relevanten Zeitraum nicht wechselten. Analog zur Auswertung des sFerritin erfolgte jeweils eine Adjustierung der ANCOVA für den entsprechenden Baseline-Wert sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren. Zur Quantifizierung des Gruppenunterschieds wird jeweils die adjustierte Mittelwertdifferenz dargestellt.</p> <p>Alle Analysen zu diesem Endpunkt basieren auf der ITT-Population, wobei nur Patient:innen mit hinreichenden Daten für die Auswertung berücksichtigt wurden, d. h. Patient:innen mit Messwerten zu den erforderlichen Zeitpunkten.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Analyse:</b> <ul style="list-style-type: none"> <li>○ Mittlere Änderung des sFerritin-Spiegels bis Woche 48</li> </ul> </li> <li>• <b>Weitere Analysen in Anhang 4G:</b> <ul style="list-style-type: none"> <li>○ Mittlere Änderung der LIC bis Woche 48</li> <li>○ Mittlere Änderung der täglichen Dosis Eisenchelator bis Woche 48</li> </ul> </li> </ul>

Studie	Operationalisierung
	ANCOVA: <i>Analysis of covariance</i> ; ITT: Intention-to-Treat; LIC: Lebereisenkonzentration ( <i>liver iron concentration</i> ); MRT: Magnetresonanztomographie; sFerritin: Serum-Ferritin

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 4-29: Bewertung des Verzerrungspotenzials für Endpunkt „Morbidity – Kontrolle der Eisenüberladung“ in der Studie BEYOND

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

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt „Kontrolle der Eisenüberladung“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung oder sonstige Aspekte, die zu einer Verzerrung führen könnten, liegen nicht vor.

Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

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.

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

### Mittlere Änderung des sFerritin-Spiegels

Tabelle 4-30: Ergebnisse für den Endpunkt „Morbidity – Kontrolle der Eisenüberladung“ – mittlere Änderung des sFerritin-Spiegels gegenüber Baseline aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel (SE) <sup>(4)</sup>	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel (SE) <sup>(4)</sup>	Mittlere Differenz [95 %-KI]; p-Wert <sup>(4)</sup>	Hedges' g <sup>(5)</sup> [95 %-KI]
<b>Änderung des sFerritin-Spiegels gegenüber Baseline</b>								
Mittlere Änderung bis Woche 48 (µg/l) <sup>(6)</sup>	92	567,77 (523,16)	84,98 (23,117)	44	528,84 (444,92)	71,48 (32,453)	13,50 [-60,96; 87,96]; 0,7205	0,06 [-0,28; 0,41]
ITT-Population; Primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Patient:innen mit mindestens einer Messung des sFerritins in den letzten 24 Wochen ihrer Behandlungsphase bis Woche 48 <sup>(3)</sup> Arithmetischer Mittelwert basierend auf den durchschnittlichen sFerritin-Werten in den 24 Wochen vor der ersten Dosis (Angabe bezieht sich auf alle randomisierten Patient:innen) <sup>(4)</sup> ANCOVA basierend auf der Änderung gegenüber Baseline, adjustiert für den Baseline-sFerritin-Wert, den Baseline-Hb-Wert und den Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche <sup>(5)</sup> Hedges' g abgeleitet aus dem Modellschätzer der mittleren Differenz <sup>(6)</sup> Durchschnitt der sFerritin-Werte in den letzten 24 Wochen der Behandlungsphase bis einschließlich Woche 48 ANCOVA: <i>Analysis of covariance</i> ; BSC: <i>Best-Supportive-Care</i> ; Hb: Hämoglobin; ITT: Intention-to-Treat; KI: Konfidenzintervall; SD: Standardabweichung; SE: Standardfehler; sFerritin: Serum-Ferritin								

Die Auswertungen zum Endpunkt „Kontrolle der Eisenüberladung“ sind klinisch relevante Endpunkte im vorliegenden Anwendungsgebiet und werden daher ergänzend berichtet, jedoch nicht zur Ableitung eines Zusatznutzens herangezogen. Zu Woche 48 kam es unter Luspatercept im Vergleich zu Placebo zu keinem statistisch signifikanten Unterschied für die mittlere Änderung des sFerritin-Spiegels (Mittlere Differenz = 13,50 µg/l; 95 %-KI: [-60,96; 87,96];  $p < 0,7205$ ; siehe mittlere Änderung des sFerritin-Spiegels Tabelle 4-30). Weitere Analysen zur Änderung der LIC und der täglichen Eisenchelator-Dosis zeigen ebenfalls keine relevanten Unterschiede (siehe Anhang 4-G).

*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.*

#### **Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

#### **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Morbidität – Kontrolle der Eisenüberladung“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### **4.3.1.3.1.4 Morbidität – Symptomatik gemäß NTDT-PRO**

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

Tabelle 4-31: Operationalisierung des Endpunkts „Morbidität – Symptomatik gemäß NTDT-PRO“

Studie	Operationalisierung
BEYOND	Das Messinstrument NTDT-PRO ist ein patientenberichteter, krankheitsspezifischer Fragebogen, der die Wahrnehmung der Patient:innen bezüglich ihrer NTDT-bedingten Symptomatik anhand der folgenden Items auf einer numerischen Rating-Skala von 0 – 10 („keine“ bis „extrem/sehr stark“) misst: <ul style="list-style-type: none"> <li>• Müdigkeit (Energienmangel) ohne körperliche Belastung</li> </ul>

Studie	Operationalisierung
	<ul style="list-style-type: none"> <li>• Müdigkeit (Energienmangel) während körperlicher Belastung</li> <li>• Schwäche (Kraftlosigkeit) ohne körperliche Belastung</li> <li>• Schwäche (Kraftlosigkeit) während körperlicher Belastung</li> <li>• Kurzatmigkeit ohne körperliche Belastung</li> <li>• Kurzatmigkeit während körperlicher Belastung</li> </ul> <p>Aus den Einzelitems werden die Gesamtscores für zwei verschiedene Domänen bestimmt, die ebenfalls eine Skalenspannweite von 0 bis 10 aufweisen:</p> <ul style="list-style-type: none"> <li>• Gesamtscore in der Domäne Müdigkeit/Schwäche (<i>tiredness/weakness</i>; T/W)</li> <li>• Gesamtscore in der Domäne Kurzatmigkeit (<i>shortness of breath</i>, SoB)</li> </ul> <p>Die Erhebung des Fragebogens erfolgte anhand eines täglich geführten Tagebuchs durchgängig ab 7 Tage vor der ersten Dosis bis Woche 24 der doppelblinden Behandlungsphase und danach über 7 Tage vor jeder zweiten Behandlungsvisite, zusätzlich vor Woche 13, 24, 37 und 48 (sofern eine Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der doppelblinden Behandlungsphase stattfand), sowie 9 Wochen nach der letzten Dosis. Für die Analyse wurden die täglichen Angaben pro Item zu einem wöchentlichen Score gemittelt. Sofern ein Item an weniger als 4 von 7 Tagen beantwortet worden war, wurde der wöchentliche Score als fehlend betrachtet.</p> <p>Der Gesamtscore in der Domäne Müdigkeit/Schwäche wird als der Mittelwert der vier Items zu Müdigkeit und Schwäche berechnet; der Gesamtscore in der Domäne Kurzatmigkeit wird als Mittelwert der beiden Items zu Kurzatmigkeit berechnet.</p> <p>Die Auswertung erfolgte jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber dem Ausgangswert. Als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung werden dabei 1,5 Punkte betrachtet (entsprechend einer Veränderung von 15 % der Skalenspannweite). Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell unter Berücksichtigung des Baseline-Werts sowie ein stratifizierter Log-Rank-Test berechnet. Die Stratifizierung erfolgte basierend auf den Stratifizierungsfaktoren der Randomisierung.</p> <p>Als Zusatzanalysen werden im Rahmen von Responderanalysen die Anteile der Patient:innen mit einer bewertungsrelevanten Verbesserung sowie einer bewertungsrelevanten Verschlechterung zu Woche 48 (und ergänzend zu Woche 24) dargestellt. Patient:innen mit einem fehlenden Wert zu Woche 48 (bzw. Woche 24) wurden von diesen Analysen ausgeschlossen. Als Schätzer des Behandlungseffekts wird das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.</p> <p>Um die mittlere Änderung der patientenberichteten Scores auch im zeitlichen Verlauf abzubilden, werden die arithmetischen Mittelwerte pro Behandlungsarm und Visite graphisch durch Verlaufskurven dargestellt.</p> <p>Weiterhin wurde für die Analyse der mittleren Änderung ein MMRM auf Basis der Daten aus dem Beobachtungszeitraum bis Woche 48 berechnet. Im entsprechenden Modell wurden, zusätzlich zur Behandlungsgruppe, der jeweilige Baseline-Score, die Visite, die Interaktion Behandlungsgruppe*Visite, die Interaktion Baseline-Score*Visite sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren als feste Effekte verwendet. Der Achsen-schnittpunkt (Intercept) und die Visite wurden zudem als zufällige Effekte definiert. Zur Quantifizierung des Behandlungseffekts wird der adjustierte Modellschätzer der Mittelwert-differenz zu Woche 48 (und ergänzend zu Woche 24) sowie die hieraus abgeleitete standardisierte Mittelwertdifferenz (SMD) in Form von Hedges' g darstellt.</p> <p>Die Analysen basieren auf der ITT-Population, wobei nur Patient:innen mit hinreichenden Daten für die jeweilige Auswertung berücksichtigt wurden, d. h. Patient:innen mit berechenbaren NTDT-PRO-Scores zu den erforderlichen Zeitpunkten. Der NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche wurden nur berechnet, wenn mindestens eines der beiden Items zu</p>

Studie	Operationalisierung
	<p>Müdigkeit und mindestens eines der beiden Items zu Schwäche beantwortet worden waren. Der NTDT-PRO-Gesamtscore in der Domäne Kurzatmigkeit wurde nur berechnet, wenn beide Items zu Kurzatmigkeit beantwortet worden waren.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores</li> <li>• <b>Zusatzanalysen:</b> <ul style="list-style-type: none"> <li>○ Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores zu Woche 24 und 48</li> <li>○ Line Plots zum Verlauf der mittleren Änderung der NTDT-PRO Gesamtscores bis zum primären Datenschnitt</li> </ul> </li> <li>• <b>Zusatzanalysen in Anhang 4G:</b> <ul style="list-style-type: none"> <li>○ Rücklaufquoten für den NTDT-PRO</li> <li>○ Mittlere Änderung der NTDT-PRO Gesamtscores (MMRM-Analyse)</li> <li>○ Auswertungen zu den NTDT-PRO Einzelitems</li> </ul> </li> </ul>
<p>HR: Hazard Ratio; ITT: Intention-to-Treat; MMRM: <i>mixed-model repeated measure</i>; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i>; SMD: standardisierte Mittelwertdifferenzen</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 4-32: Bewertung des Verzerrungspotenzials für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ in der Studie BEYOND

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
BEYOND	niedrig	ja	ja	ja	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt „Symptomatik gemäß NTDT-PRO“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor.

Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert (siehe Tabelle 4-33).

Tabelle 4-33: Beobachtungszeiten für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ in der Studie BEYOND

<b>Studie BEYOND Beobachtungszeit in Wochen</b>	<b>Luspatercept + BSC <sup>(1)</sup> (N=96)</b>	<b>Placebo + BSC <sup>(1)</sup> (N=49)</b>
Mittelwert (SD)	73,3 (35,26)	59,8 (31,69)
Median (Min; Max)	79,4 (0; 130,3)	48,9 (3,0; 116,0)
Primärer Datenschnitt: 14.09.2020; ITT-Population <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie. BSC: <i>Best-Supportive-Care</i> ; ITT: <i>Intention-to-Treat</i> ; Max: <i>Maximum</i> ; Min: <i>Minimum</i> ; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> ; SD: <i>Standardabweichung</i>		

Bei Patient:innen, die die Therapie abbrachen, wurde der NTDT-PRO weiterhin bis mindestens Woche 48 sowie 9 Wochen nach der letzten Dosis erhoben.

Während zu Woche 24 hinreichend hohe Rücklaufquoten vorlagen (80,2 % im Luspatercept-Arm bzw. 79,6 % Placebo-Arm), sanken die Rücklaufquoten zu Woche 48 zwar unter die Marke von 70 % bezogen auf die ITT-Population (50,0 % bzw. 55,1 %, vgl. Anhang 4-G), die Effekte der Responderanalysen zu Woche 24 und 48 blieben jedoch konsistent.

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

*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.*

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

**Hauptanalyse: Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores**

Tabelle 4-34: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
<b>Zeit bis zur ersten bewertungsrelevanten Verbesserung</b>								
Gesamtscore Domäne Müdigkeit/Schwäche	94	45 (47,9)	N.A. [3,5; N.A.]	49	22 (44,9)	20,3 [1,6; N.A.]	1,03 [0,61; 1,71]	0,8972
Gesamtscore Domäne Kurzatmigkeit	94	40 (42,6)	N.A. [7,7; N.A.]	49	17 (34,7)	N.A. [11,9; N.A.]	1,20 [0,67; 2,14]	0,2348
<b>Zeit bis zur ersten bewertungsrelevanten Verschlechterung</b>								
Gesamtscore Domäne Müdigkeit/Schwäche	94	19 (20,2)	N.A. [N.A.; N.A.]	49	10 (20,4)	N.A. [N.A.; N.A.]	0,94 [0,43; 2,04]	0,8670
Gesamtscore Domäne Kurzatmigkeit	94	23 (24,5)	N.A. [N.A.; N.A.]	49	12 (24,5)	N.A. [N.A.; N.A.]	1,00 [0,49; 2,01]	0,9714
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.								
<sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und mindestens einer Post-Baseline Bewertung im Studienverlauf								
<sup>(3)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verbesserung/Verschlechterung								
<sup>(4)</sup> Cox-Regressionsmodell adjustiert für den Ausgangswert und stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(5)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i>								

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

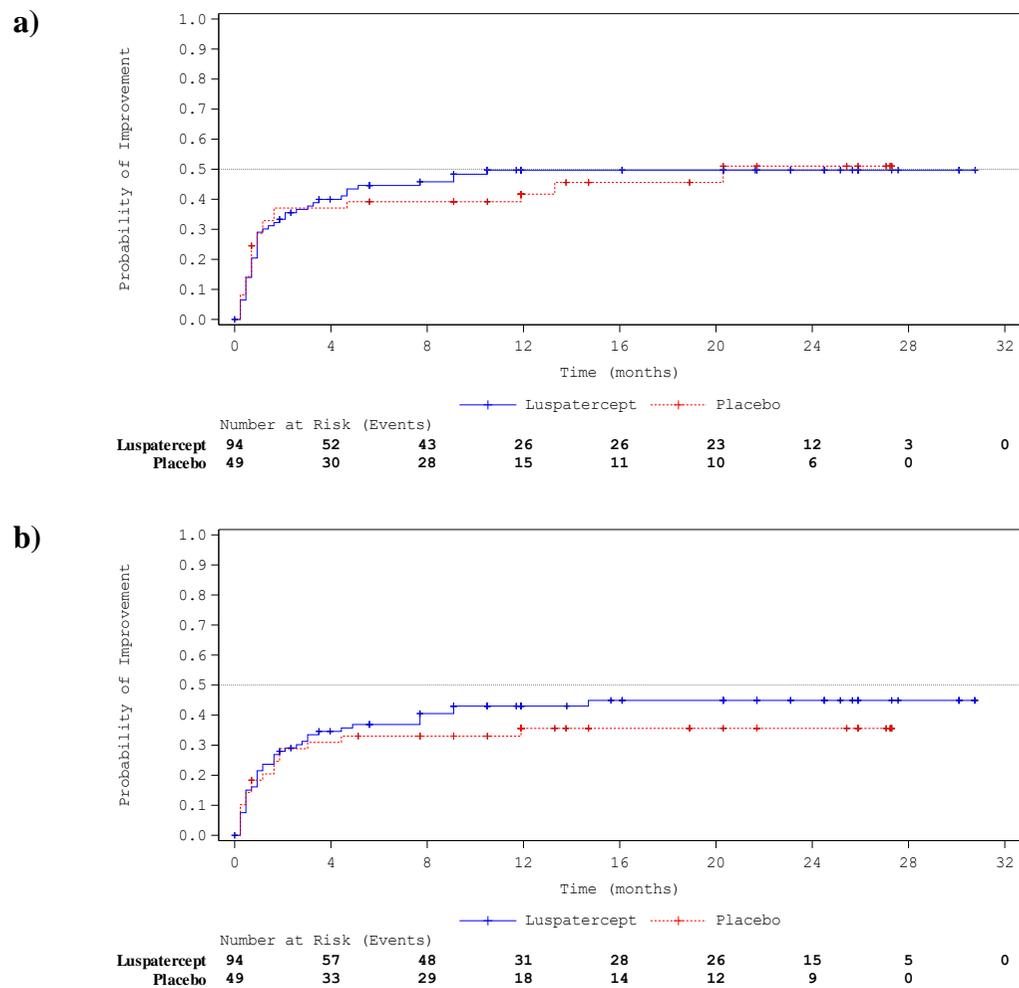


Abbildung 4-5: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung des NTDT-PRO Gesamtscores in der Domäne a) Müdigkeit/Schwäche und b) Kurzatmigkeit

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

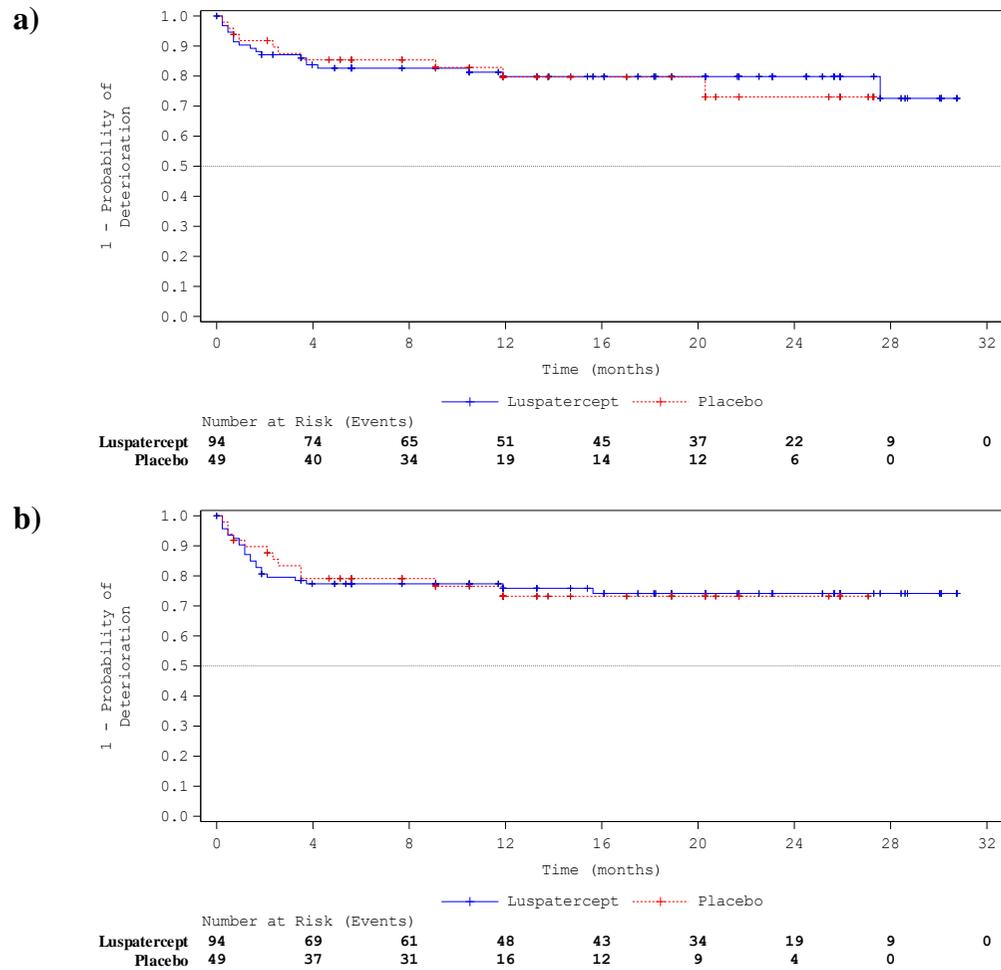


Abbildung 4-6: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne a) Müdigkeit/Schwäche und b) Kurzatmigkeit

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

Für den Endpunkt „Symptomatik gemäß NTDT-PRO“ zeigt sich in der Studie BEYOND kein statistisch signifikanter Unterschied in den Ereigniszeitanalysen zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores (siehe Tabelle 4-34 und Abbildung 4-5 bzw. Abbildung 4-6).

Weitere Zusatzanalysen sind im Folgenden dargestellt.

### Zusatzanalysen

#### *Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores zu Woche 24 und 48*

Tabelle 4-35: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der NTDT-PRO Gesamtscores zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verbesserung des NTDT-PRO Gesamtscores in der Domäne Müdigkeit/Schwäche</b>							
Verbesserung zu Woche 24	76	27 (35,5)	39	7 (17,9)	0,19 [0,04; 0,35]	2,91 [1,10; 7,71]	2,06 [1,02; 4,17]; <b>0,0434</b>
Verbesserung zu Woche 48	48	20 (41,7)	27	3 (11,1)	0,28 [0,10; 0,46]	5,93 [1,43; 24,56]	3,67 [1,12; 11,99]; <b>0,0317</b>
<b>Bewertungsrelevante Verbesserung des NTDT-PRO Gesamtscores in der Domäne Kurzatmigkeit</b>							
Verbesserung zu Woche 24	76	21 (27,6)	39	4 (10,3)	0,19 [0,05; 0,33]	3,79 [1,20; 11,94]	2,87 [1,09; 7,59]; <b>0,0332</b>
Verbesserung zu Woche 48	48	17 (35,4)	27	3 (11,1)	0,23 [0,06; 0,40]	5,12 [1,13; 23,13]	3,56 [0,95; 13,33]; 0,0594

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verschlechterung des NTDT-PRO Gesamtscores in der Domäne Müdigkeit/Schwäche</b>							
Verschlechterung zu Woche 24	76	6 (7,9)	39	3 (7,7)	0,00 [-0,10; 0,11]	1,05 [0,26; 4,29]	1,04 [0,28; 3,96]; 0,9494
Verschlechterung zu Woche 48	48	4 (8,3)	27	3 (11,1)	-0,02 [-0,16; 0,12]	0,77 [0,15; 3,99]	0,80 [0,20; 3,26]; 0,7529
<b>Bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne Kurzatmigkeit</b>							
Verschlechterung zu Woche 24	76	7 (9,2)	39	3 (7,7)	0,01 [-0,10; 0,12]	1,18 [0,30; 4,70]	1,17 [0,32; 4,22]; 0,8124
Verschlechterung zu Woche 48	48	4 (8,3)	27	6 (22,2)	-0,13 [-0,30; 0,04]	0,30 [0,07; 1,35]	0,40 [0,12; 1,28]; 0,1223
ITT-Population; Primärer Datenschnitt vom 14.09.2020							
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.							
<sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und einer Bewertung zu Woche 24 bzw. 48							
<sup>(3)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche							
<sup>(4)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet.							
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> ; OR: Odds Ratio; RD: Risiko Differenz; RR: Relatives Risiko							

In Bezug auf den Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung der NTDT-PRO Gesamtscores zu Woche 24 und 48 zeigten sich in der Studie BEYOND statistisch signifikante Effekte zugunsten von Luspatercept (siehe Tabelle 4-35).

In Bezug auf den NTDT-PRO Gesamtscore in der Domäne Müdigkeit/Schwäche betrug der Anteil der Patient:innen, die zu Woche 24 eine bewertungsrelevante Verbesserung erreichten, 35,5 % im Luspatercept-Arm gegenüber 17,9 % im Placebo-Arm (p = 0,0434). Dieser signifikante Vorteil setzte sich zu Woche 48 mit 41,7 % gegenüber 11,1 % fort (p = 0,0317).

---

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

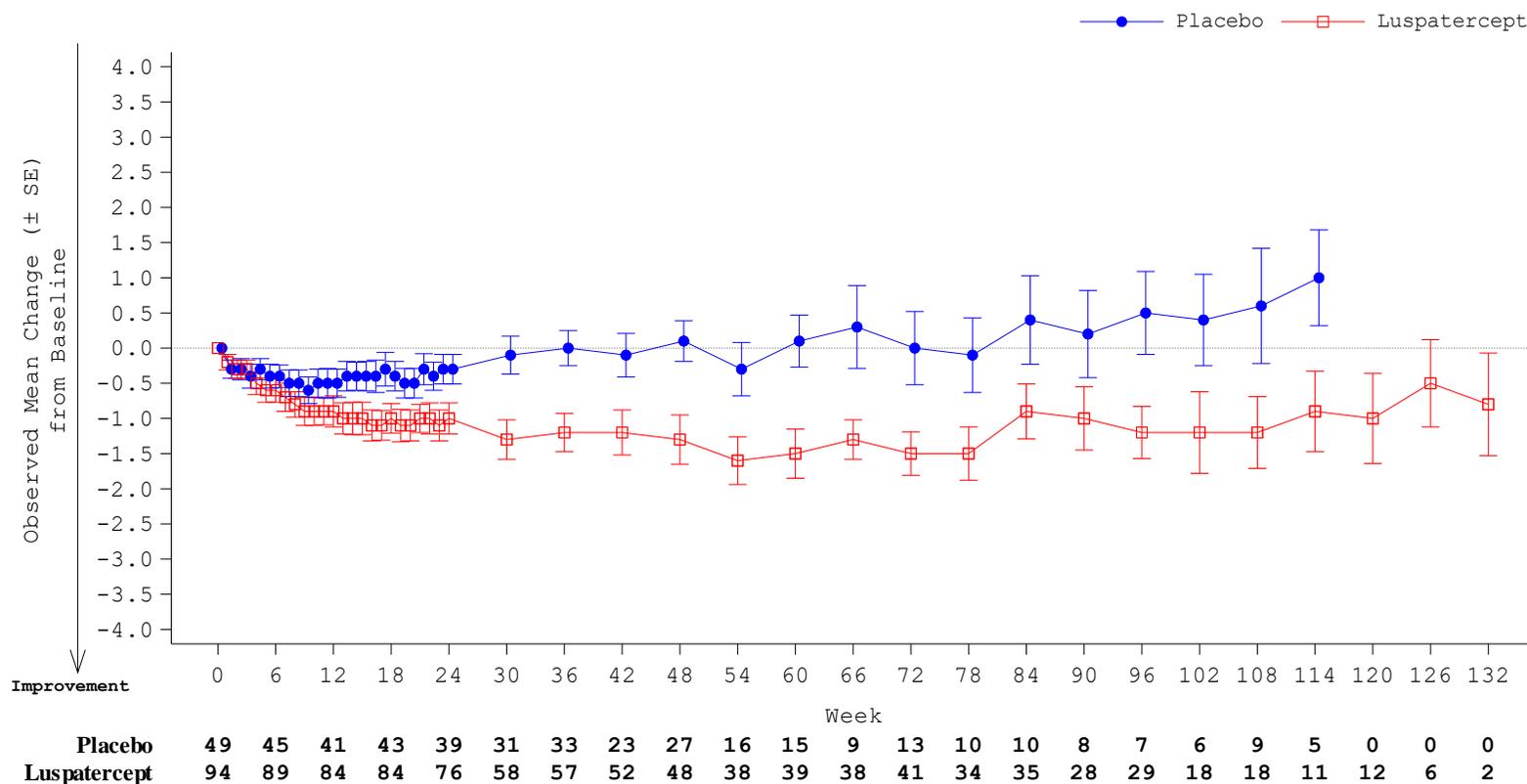
In Bezug auf den NTDT-PRO Gesamtscore in der Domäne Kurzatmigkeit betrug der Anteil der Patient:innen, die zu Woche 24 eine bewertungsrelevante Verbesserung erreichten, 27,6 % im Luspatercept-Arm gegenüber 10,3 % im Placebo-Arm ( $p = 0,0332$ ). Dieser Vorteil blieb zu Woche 48 mit 35,4 % gegenüber 11,1 % der Patient:innen numerisch erhalten ( $p = 0,0594$ ).

Die Analysen zum Anteil der Patient:innen mit einer bewertungsrelevanten Verschlechterung zeigen jeweils keine signifikanten Unterschiede.

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

**Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß NTDT-PRO**

a)



b)

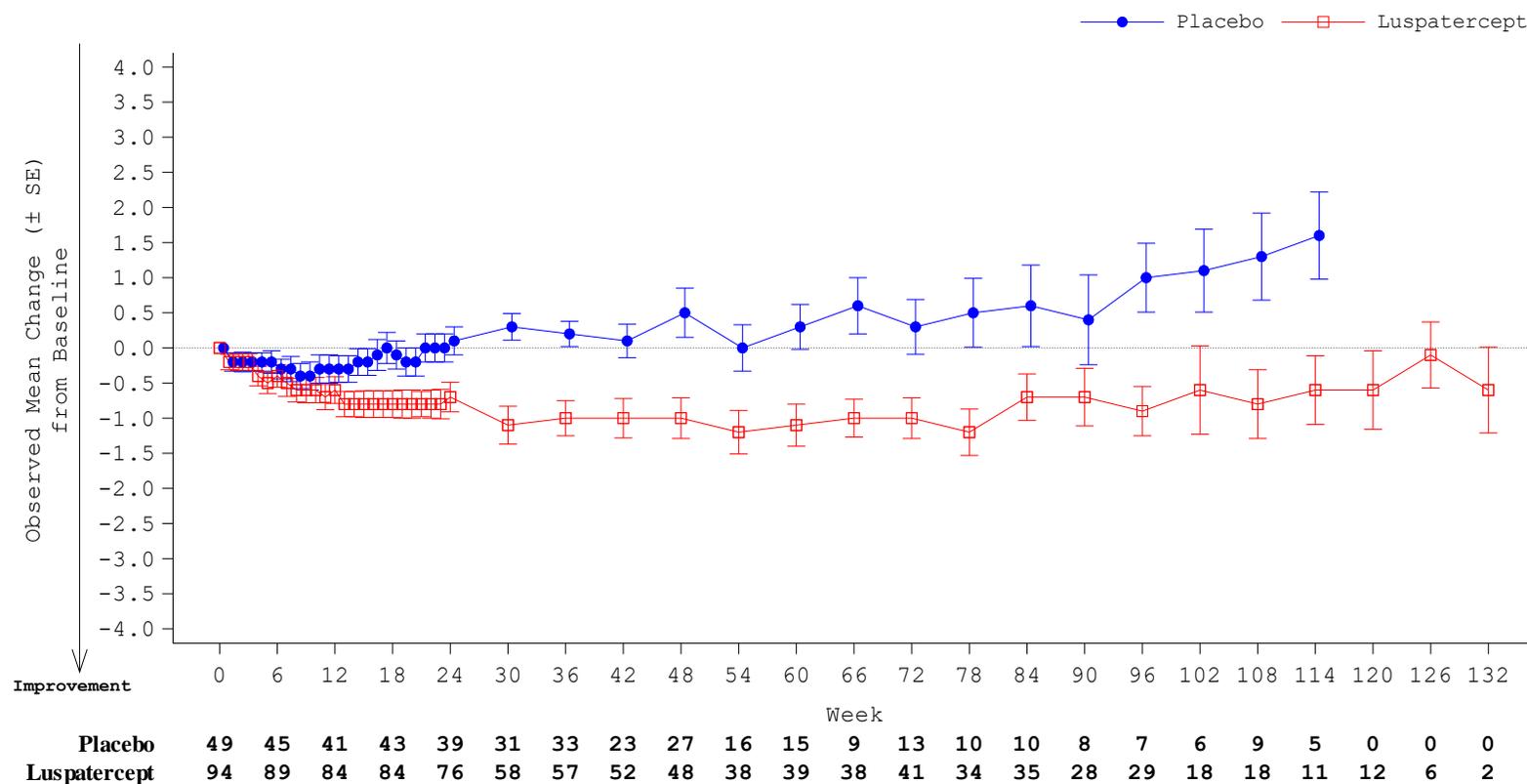


Abbildung 4-7: Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß NTDT-PRO Gesamtscore in der Domäne a) Müdigkeit/Schwäche und b) Kurzatmigkeit

Die Vorteile von Luspatercept in den Responderanalysen zur Verbesserung der Symptomatik gemäß NTDT-PRO sind auch im Studienverlauf der mittleren Änderung der NTDT-PRO Gesamtscores deutlich erkennbar (siehe Abbildung 4-7).

Darüber hinaus zeigten sich auch bei der MMRM-Analyse zur mittleren Änderung gegenüber dem Ausgangswert in beiden NTDT-PRO Gesamtscores sowohl zu Woche 24 als auch zu Woche 48 jeweils ein signifikanter Vorteil für den Luspatercept-Arm (siehe Anhang 4-G).

**Zusammenfassung der Ergebnisse zum Endpunkt Symptomatik gemäß NTDT-PRO**

Zwar ergaben sich bei den Auswertungen zur ersten bewertungsrelevanten Verbesserung und Verschlechterung keine signifikanten Vorteile für Luspatercept, jedoch zeigen die zusätzlichen Responderanalysen hinsichtlich einer bewertungsrelevanten Verbesserung für beide NTDT-PRO Gesamtscores statistisch signifikante Vorteile. Entsprechende Unterschiede sind auch in den Abbildungen der mittleren Verläufe und den MMRM-Analysen ersichtlich. Zudem lassen die Abbildungen vermuten, dass die Vorteile im Rahmen der Ereigniszeitanalysen aufgrund von vorübergehenden Anfangsverbesserungen im Placebo-Arm ggf. unzureichend abgebildet werden.

Insgesamt sind die Ergebnisse zum NTDT-PRO positiv zu bewerten. Die Zusatzanalysen deuten auf eine spürbare Verbesserung der krankheitsspezifischen Symptomatik unter Luspatercept hin.

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

**Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

**Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Morbidität – Symptomatik gemäß NTDT-PRO“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### 4.3.1.3.1.5 Morbidität – Symptomatik gemäß PGI-S und PGI-C

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

Tabelle 4-36: Operationalisierung des Endpunkts „Morbidität – Symptomatik gemäß PGI-S und PGI-C“

Studie	Operationalisierung
BEYOND	<p>Die Messinstrumente Patient Global Impression of Severity (PGI-S) und Patient Global Impression of Change (PGI-C) sind patientenberichtete Fragebögen, die die Wahrnehmung der Patient:innen bezüglich ihrer Thalassämie-bedingten Symptomatik messen.</p> <p>Der PGI-S besteht aus einer einzigen Frage, die die Patient:innen bittet, die Schwere ihrer Thalassämie-bedingten Symptomatik anhand einer 11 Punkte umfassenden numerischen Rating-Skala (von 0 = „keine Symptome“ bis 10 = „sehr schwere Symptome“) zu bewerten. Auch der PGI-C besteht aus einer einzigen Frage, wobei die Patient:innen hier angehalten werden auf einer 7 Punkte umfassenden verbalen Rating-Skala (von „sehr viel besser“ bis „sehr viel schlechter“) die gesamte Änderung ihrer Thalassämie-bedingten Symptomatik seit Beginn der Studie einzuschätzen.</p> <p>Die Erhebung des PGI-S erfolgte anhand eines täglich geführten Tagebuchs (gemeinsam mit dem NTDT-PRO) durchgängig ab 7 Tage vor der ersten Dosis bis Woche 24 der doppelblinden Behandlungsphase und danach über 7 Tage vor jeder zweiten Behandlungsvisite, zusätzlich vor Woche 13, 24, 37 und 48 (sofern eine Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der doppelblinden Behandlungsphase stattfand), sowie 9 Wochen nach der letzten Dosis. Die Erhebung des PGI-C erfolgte bei jeder zweiten Behandlungsvisite während der doppelblinden Behandlungsphase (beginnend mit der dritten Dosis), zusätzlich zu Woche 13, 24, 37 und 48 (sofern eine Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der doppelblinden Behandlungsphase stattfand), sowie 9 Wochen nach der letzten Dosis.</p> <p>Die Auswertung erfolgte jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber Baseline. Beim PGI-S wurden 1,5 Punkte (entsprechend 15 % der Skalenspannweite) als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung herangezogen. Beim PGI-C wurden die beiden besten Bewertungen der Veränderung („viel besser“ und „sehr viel besser“) als bewertungsrelevante Verbesserung und die beiden schlechtesten Bewertungen („viel schlechter“ und „sehr viel schlechter“) als bewertungsrelevante Verschlechterung betrachtet. Als alternativer Schwellenwert wird im Rahmen einer Zusatzanalyse die Auswertung des PGI-C bzgl. jeglicher Verbesserung/Verschlechterung der Symptomatik betrachtet (d. h. als Verbesserung zählen die Antworten „ein wenig besser“, „viel besser“ und „sehr viel besser“, als Verschlechterung die Antworten „ein wenig schlechter“, „viel schlechter“ und „sehr viel schlechter“).</p> <p>Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Beim PGI-S wurde der Baseline-Score als Kovariate im Cox-Modell berücksichtigt. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert.</p> <p>Als Zusatzanalysen werden im Rahmen von Responderanalysen die Anteile der Patient:innen mit einer bewertungsrelevanten Verbesserung sowie einer bewertungsrelevanten Verschlechterung zu Woche 48 (und ergänzend zu Woche 24) dargestellt. Patient:innen mit einem fehlenden Wert zu Woche 48 (bzw. Woche 24) wurden von diesen Analysen ausgeschlossen. Als Schätzer des Behandlungseffekts wird das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.</p> <p>Um die Veränderung der patientenberichteten Scores auch im zeitlichen Verlauf graphisch abzubilden, werden die arithmetischen Mittelwerte pro Behandlungsarm und Visite als Verlaufskurve (PGI-S) bzw. als Balkendiagramm der Verteilung der Bewertungen pro Visite (PGI-C)</p>

Studie	Operationalisierung
	<p>dargestellt.</p> <p>Weiterhin wurde für die Analyse der mittleren Änderung des PGI-S gegenüber dem Ausgangswert als stetige Variable ein MMRM auf Basis der Daten aus dem Beobachtungszeitraum bis Woche 48 berechnet. Im entsprechenden Modell wurden, zusätzlich zur Behandlungsgruppe, der jeweilige Baseline-Score, die Visite, die Interaktion Behandlungsgruppe*Visite, die Interaktion Baseline-Score*Visite sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren als feste Effekte verwendet. Der Achsenschnittpunkt (Intercept) und die Visite wurden zudem als zufällige Effekte definiert. Zur Quantifizierung des Behandlungseffekts wird der adjustierte Modellschätzer der Mittelwertdifferenz zu Woche 48 (und ergänzend zu Woche 24) sowie die hieraus abgeleitete standardisierte Mittelwertdifferenz (SMD) in Form von Hedges' g darstellt.</p> <p>Alle Analysen zu diesem Endpunkt basieren auf der ITT-Population, wobei nur Patient:innen mit hinreichenden Daten für die Auswertung berücksichtigt wurden, d. h. Patient:innen mit Bewertungen des PGI-S bzw. PGI-C zu den erforderlichen Zeitpunkten.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung des PGI-S und PGI-C</li> <li>• <b>Zusatzanalysen:</b> <ul style="list-style-type: none"> <li>○ Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des PGI-S und PGI-C zu Woche 24 und 48</li> <li>○ Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß PGI-S bis zum primären Datenschnitt</li> <li>○ Bar Chart zur Verteilung der Änderung der Symptomatik gemäß PGI-C pro Visite bis zum primären Datenschnitt</li> </ul> </li> <li>• <b>Zusatzanalysen in Anhang 4G:</b> <ul style="list-style-type: none"> <li>○ Rücklaufquoten zum PGI-S und PGI-C</li> <li>○ Zeit bis zur ersten Verbesserung bzw. Verschlechterung des PGI-C</li> <li>○ Mittlere Änderung des PGI-S (MMRM-Analyse)</li> </ul> </li> </ul>
<p>HR: Hazard Ratio; MMRM: <i>mixed-model repeated measure</i>; PGI-C: <i>patient global impression of change</i>; PGI-S: <i>patient global impression of severity</i>; SMD: standardisierte Mittelwertdifferenzen</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 4-37: Bewertung des Verzerrungspotenzials für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ in der Studie BEYOND

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

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt „Symptomatik gemäß PGI-S und PGI-C“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert (siehe Tabelle 4-38).

Tabelle 4-38: Beobachtungszeiten für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ in der Studie BEYOND

Studie BEYOND Beobachtungszeit in Wochen	Luspatercept + BSC <sup>(1)</sup> (N=96)	Placebo + BSC <sup>(1)</sup> (N=49)
PGI-S		
Mittelwert (SD)	73,3 (35,26)	59,8 (31,69)
Median (Min; Max)	79,4 (0; 130,3)	48,9 (3,0; 116,0)
PGI-C		
Mittelwert (SD)	89,9 (30,30)	70,1 (28,32)
Median (Min; Max)	95,9 (12,4; 130,4)	60,6 (19,4; 120,1)
Primärer Datenschnitt: 14.09.2020; ITT-Population		
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.		
BSC: <i>Best-Supportive-Care</i> ; ITT: <i>Intention-to-Treat</i> ; Max: <i>Maximum</i> ; Min: <i>Minimum</i> ; PGI-C: <i>patient global impression of change</i> ; PGI-S: <i>patient global impression of severity</i> ; SD: <i>Standardabweichung</i>		

Bei Patient:innen, die die Therapie abbrechen, wurden der PGI-S und PGI-C weiterhin bis mindestens Woche 48 sowie 9 Wochen nach der letzten Dosis erhoben.

Für den PGI-C ergaben sich sowohl zu Woche 24, als auch zu Woche 48 hohe Rücklaufquoten (über 70 % in beiden Armen, vgl. Anhang 4-G). Für den PGI-S waren die Rücklaufquoten zu Woche 24 ebenfalls hinreichend hoch (80,2 % im Luspatercept-Arm bzw. 79,6 % im Placebo-Arm); zwar sanken die Rücklaufquoten für den PGI-S zu Woche 48 unter die Marke von 70 %

bezogen auf die ITT-Population (50,0 % bzw. 55,1 %), jedoch waren die Ergebnisse der Responderanalysen zu Woche 24 und 48 konsistent.

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder die Fragebögen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. In Anbetracht der Größe der beobachtenden Effekte der Ereigniszeitanalysen sowie der konsistenten Responderanalysen, ist jedoch davon auszugehen, dass dies im Ergebnis keine entscheidende Rolle spielt.

Insgesamt wird das Verzerrungspotenzial beim PGI-S und PGI-C vor diesem Hintergrund als niedrig eingestuft.

*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.*

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

**Hauptanalyse: Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung des PGI-S und PGI-C**

Tabelle 4-39: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
<b>Zeit bis zur ersten bewertungsrelevanten Verbesserung</b>								
PGI-S	94	39 (41,5)	N.A. [4,4; N.A.]	49	11 (22,4)	N.A. [N.A.; N.A.]	2,28 [1,15; 4,50]	<b>0,0244</b>
PGI-C	94	70 (74,5)	4,4 [3,9; 5,7]	48	9 (18,8)	N.A. [N.A.; N.A.]	6,44 [3,19; 12,97]	<b>&lt; 0,0001</b>
<b>Zeit bis zur ersten bewertungsrelevanten Verschlechterung</b>								
PGI-S	94	25 (26,6)	N.A. [27,6; N.A.]	49	17 (34,7)	N.A. [14,7; N.A.]	0,71 [0,38; 1,32]	0,2191
PGI-C	94	6 (6,4)	N.A. [N.A.; N.A.]	48	5 (10,4)	N.A. [N.A.; N.A.]	0,51 [0,15; 1,71]	0,2686
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.								
<sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline (mit Ausnahme des PGI-C) und mindestens einer Post-Baseline Bewertung im Studienverlauf								
<sup>(3)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verbesserung/Verschlechterung								
<sup>(4)</sup> Cox-Regressionsmodell stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche; beim PGI-S wurde das Modell zudem für den Ausgangswert adjustiert								
<sup>(5)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

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

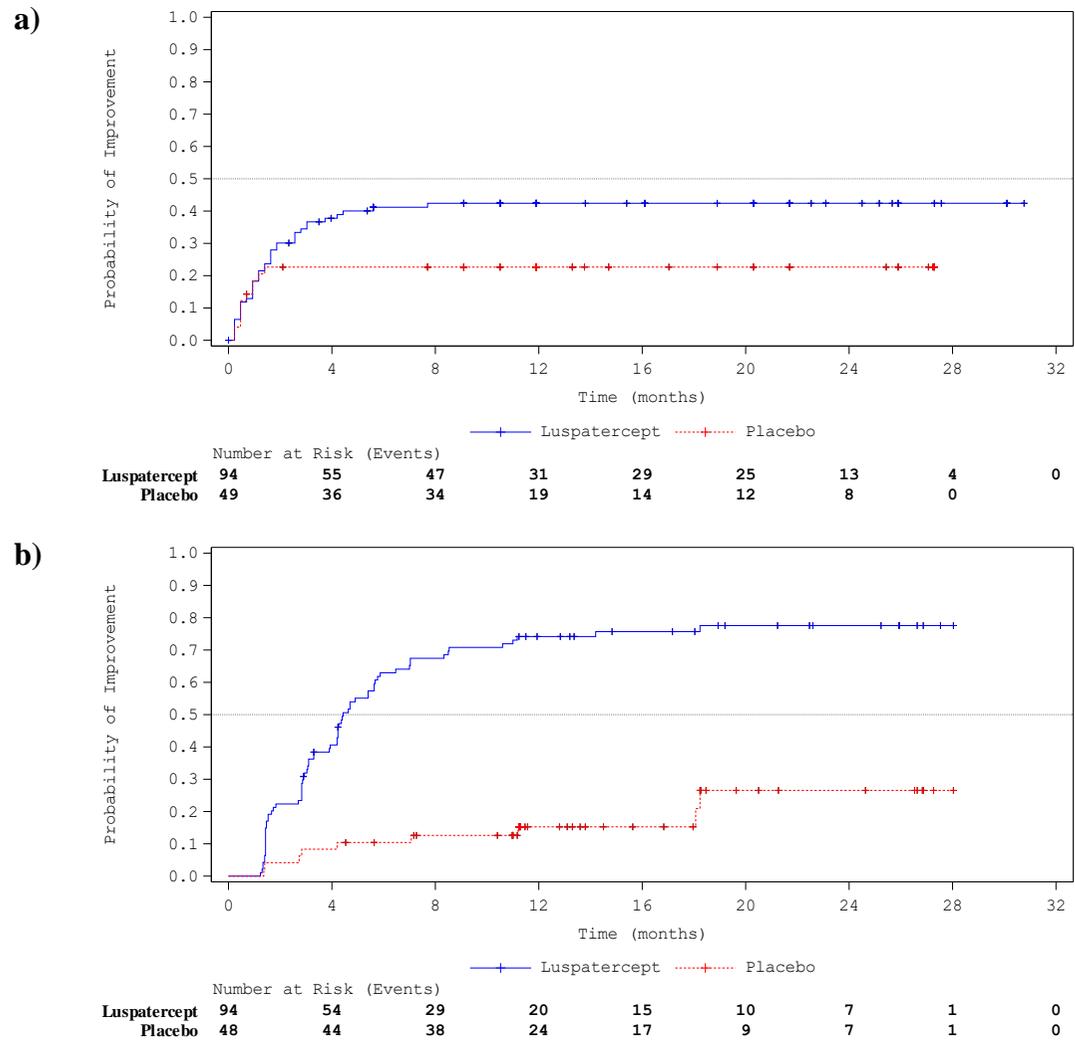


Abbildung 4-8: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung a) des PGI-S und b) des PGI-C

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

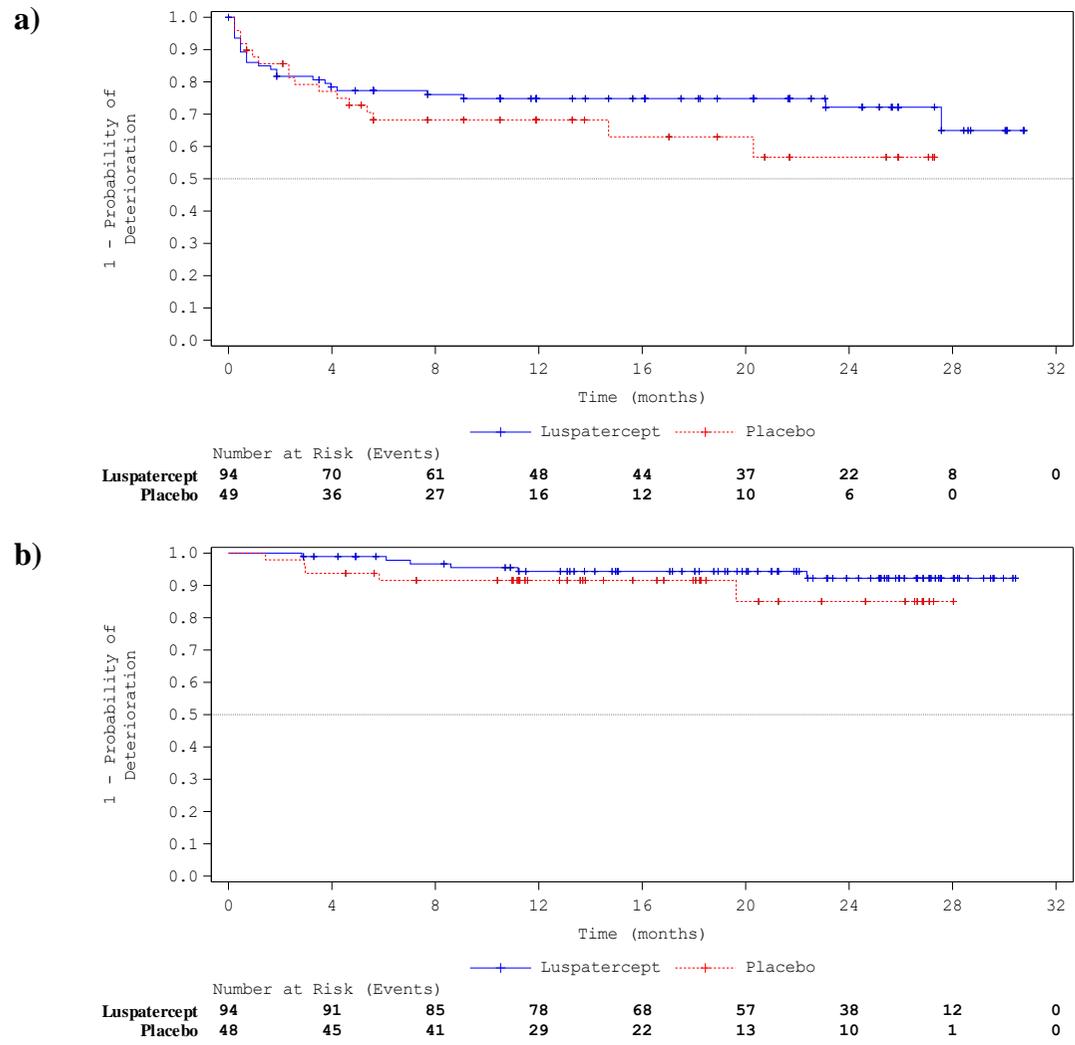


Abbildung 4-9: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung a) des PGI-S und b) des PGI-C

Für den Endpunkt „Symptomatik gemäß PGI-S und PGI-C“ zeigten sich in der Studie BEYOND statistisch signifikante Effekte zugunsten von Luspatercept.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des PGI-S ergab sich ein signifikanter Effekt mit einem HR von 2,28 (95 %-KI [1,15; 4,50];  $p = 0,0244$ ), d. h. die Chance auf eine bewertungsrelevante Verbesserung des PGI-S war unter Luspatercept mehr als doppelt so hoch wie unter Placebo. Wie in Abbildung 4-8a) ersichtlich, wurde der Median der Zeit bis zur ersten bewertungsrelevanten Verbesserung in beiden Behandlungsarmen zum Zeitpunkt des Datenschnitts noch nicht erreicht.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des PGI-C ergab sich ebenfalls ein signifikanter Effekt mit einem HR von 6,44 (95 %-KI [3,19; 12,97];  $p < 0,0001$ ), d. h. die Chance auf eine bewertungsrelevante Verbesserung des PGI-C war unter Luspatercept mehr als 6-mal höher als unter Placebo. Wie in Abbildung 4-8b) ersichtlich, betrug die mediane Zeit bis zur ersten bewertungsrelevanten Verbesserung im Luspatercept-Arm 4,4 Monate (95 %-KI [3,9; 5,7]); im Placebo-Arm wurde der Median der Zeit bis zur ersten bewertungsrelevanten Verbesserung nicht erreicht. Bei der Zusatzanalyse unter Berücksichtigung jeglicher Verbesserungen (einschließlich Verbesserungen, die als „ein wenig besser“ bewertet wurden), zeigte sich ein konsistentes Ergebnis (siehe Anhang 4-G).

Bei den Ereigniszeitanalysen zur ersten bewertungsrelevanten Verschlechterung zeigten sich keine signifikanten Unterschiede.

Weitere Zusatzanalysen sind im Folgenden dargestellt.

**Zusatzanalysen****Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des PGI-S und PGI-C zu Woche 24 und 48**

Tabelle 4-40: Ergebnisse für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des PGI-S und PGI-C zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verbesserung des PGI-S</b>							
Verbesserung zu Woche 24	76	23 (30,3)	39	4 (10,3)	0,22 [0,08; 0,35]	4,38 [1,38; 13,87]	3,08 [1,19; 7,95]; <b>0,0202</b>
Verbesserung zu Woche 48	48	18 (37,5)	27	3 (11,1)	0,20 [0,03; 0,38]	4,03 [1,00; 16,24]	2,71 [0,88; 8,33]; 0,0827
<b>Bewertungsrelevante Verbesserung des PGI-C</b>							
Verbesserung zu Woche 24	75	39 (52,0)	41	3 (7,3)	0,45 [0,31; 0,59]	14,19 [3,90; 51,56]	7,38 [2,37; 23,02]; <b>0,0006</b>
Verbesserung zu Woche 48	73	38 (52,1)	40	3 (7,5)	0,45 [0,31; 0,59]	13,45 [3,71; 48,72]	7,08 [2,29; 21,87]; <b>0,0007</b>
<b>Bewertungsrelevante Verschlechterung des PGI-S</b>							
Verschlechterung zu Woche 24	76	8 (10,5)	39	6 (15,4)	-0,05 [-0,18; 0,09]	0,66 [0,21; 2,06]	0,70 [0,26; 1,86]; 0,4694
Verschlechterung zu Woche 48	48	6 (12,5)	27	5 (18,5)	-0,06 [-0,23; 0,12]	0,63 [0,17; 2,43]	0,70 [0,25; 1,98]; 0,5012

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verschlechterung des PGI-C</b>							
Verschlechterung zu Woche 24	75	1 (1,3)	41	2 (4,9)	-0,03 [-0,10; 0,04]	0,32 [0,03; 3,56]	0,32 [0,03; 3,25]; 0,3345
Verschlechterung zu Woche 48	73	1 (1,4)	40	0	0,01 [-0,01; 0,04]	N.A.	N.A.
ITT-Population; Primärer Datenschnitt vom 14.09.2020							
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.							
<sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline (mit Ausnahme des PGI-C) und einer Bewertung zu Woche 24 bzw. 48							
<sup>(3)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche							
<sup>(4)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet.							
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; OR: Odds Ratio; RD: Risiko Differenz; RR: Relatives Risiko							

Die Ergebnisse zum Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung oder Verschlechterung zu Woche 24 und 48 stützen die Ergebnisse der Ereigniszeitanalyse (siehe Tabelle 4-40).

Im PGI-S erreichten zu Woche 24 im Luspatercept-Arm 30,3 % der Patient:innen und im Placebo-Arm 10,3 % der Patient:innen eine bewertungsrelevante Verbesserung ( $p = 0,0202$ ). Dieser Vorteil bleibt zu Woche 48 mit 37,5 % der Patient:innen im Luspatercept-Arm und 11,1 % der Patient:innen im Placebo-Arm numerisch erhalten ( $p = 0,0827$ ).

Im PGI-C erreichten zu Woche 24 im Luspatercept-Arm 52,0 % der Patient:innen und im Placebo-Arm 7,3 % der Patient:innen eine bewertungsrelevante Verbesserung ( $p = 0,0006$ ). Dieser signifikante Vorteil setzte sich zu Woche 48 mit einer bewertungsrelevanten Verbesserung bei 52,1 % der Patient:innen im Luspatercept-Arm und 7,5 % der Patient:innen im Placebo-Arm fort ( $p = 0,0007$ ).

Bei den Analysen zum Anteil der Patient:innen mit einer bewertungsrelevanten Verschlechterung zeigten sich analog zur Ereigniszeitanalyse keine signifikanten Unterschiede.

**Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß PGI-S**

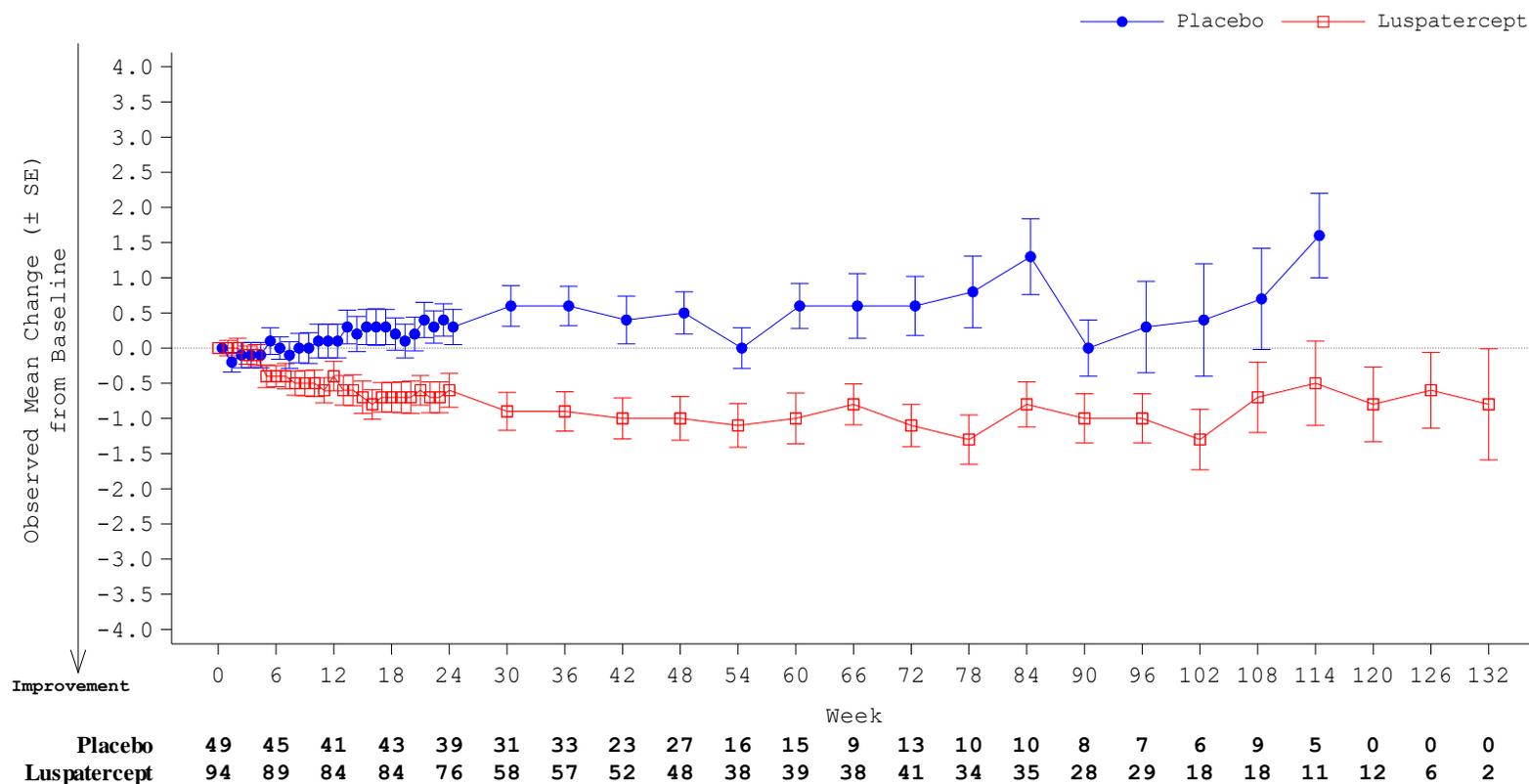
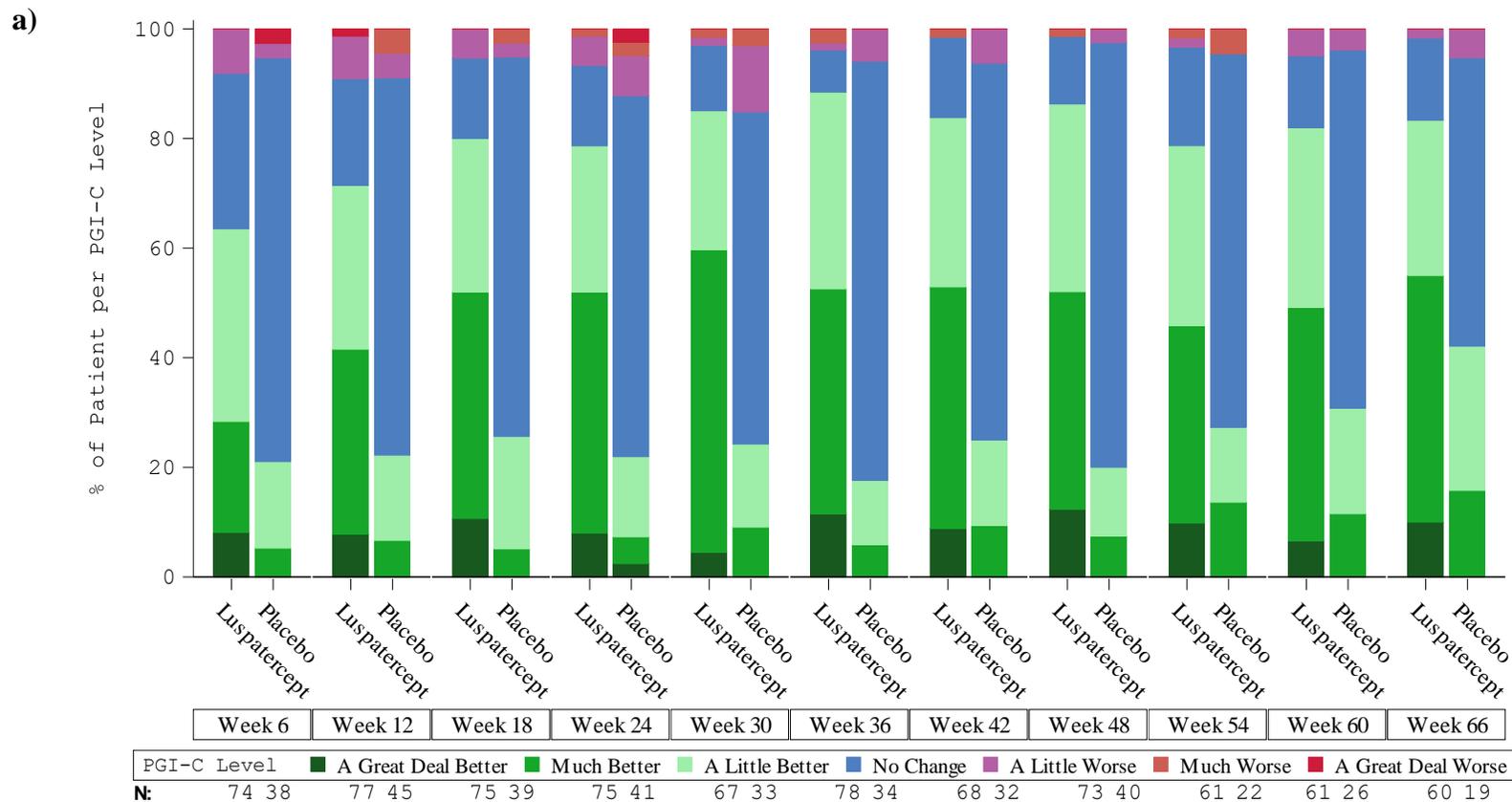


Abbildung 4-10: Line Plot zum Verlauf der mittleren Änderung der Symptomatik gemäß PGI-S

Die Vorteile von Luspatercept hinsichtlich der Verbesserung der Symptomatik gemäß PGI-S sind auch im Verlauf der mittleren Änderungen deutlich erkennbar (siehe Abbildung 4-10).

Darüber hinaus zeigte sich auch bei der MMRM-Analyse zur mittleren Änderung des PGI-S gegenüber dem Ausgangswert sowohl zu Woche 24 als auch zu Woche 48 jeweils ein signifikanter Vorteil für den Luspatercept-Arm (siehe Anhang 4-G).

**Bar Chart zur Verteilung der Änderung der Symptomatik gemäß PGI-C pro Visite**



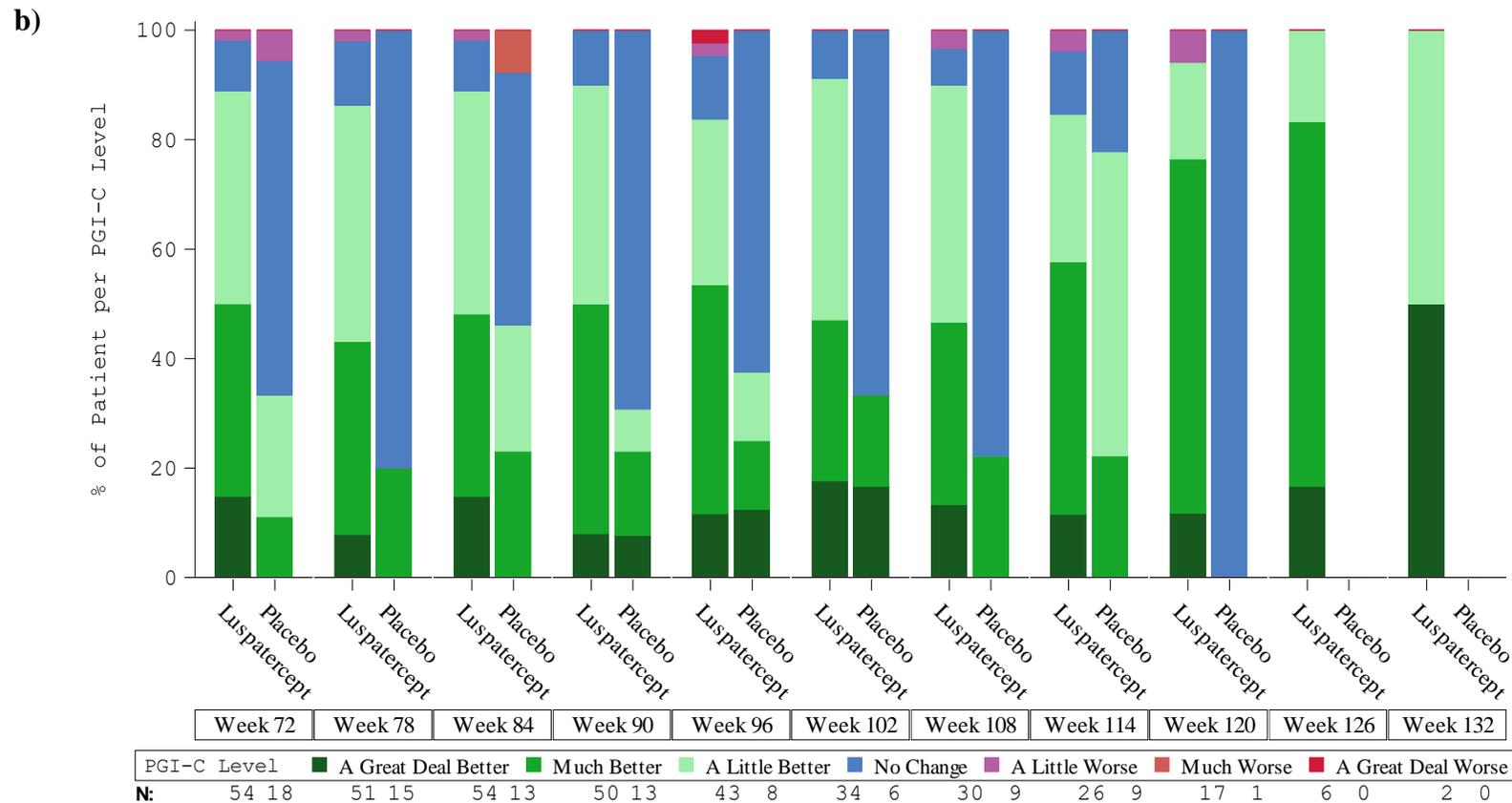


Abbildung 4-11: Bar Chart zur Änderung der Symptomatik gemäß PGI-C für a) Woche 6 bis Woche 66 und b) Woche 72 bis Woche 132

Die Vorteile von Luspaterecept in der Verbesserung der Symptomatik gemäß PGI-C sind auch im Studienverlauf anhand der Verteilung der Bewertungen pro Visite deutlich erkennbar (siehe Abbildung 4-11).

## **Zusammenfassung der Ergebnisse zum Endpunkt Symptomatik gemäß PGI-S und PGI-C**

Sowohl beim PGI-S als auch beim PGI-C zeigte sich in Bezug auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung jeweils ein sehr deutlicher, statistisch signifikanter Vorteil für Luspatercept im Vergleich zu Placebo. Diese Ergebnisse wurden durch die Zusatzanalysen zum Anteil der Patient:innen mit einer signifikanten Verbesserung zu Woche 24 und 48 bestätigt und spiegeln sich darüber hinaus auch eindeutig in den Ergebnissen der MMRM-Analysen und den Verlaufskurven zum PGI-S bzw. in der Abbildung der PGI-C Bewertungen im Studienverlauf wider.

Die Ergebnisse zum vorliegenden Endpunkt verdeutlichen, dass sich die Steigerung und langfristige Stabilisierung des Hb-Werts und einer anhaltenden Linderung der Anämie durch Luspatercept bei NTDT-Patient:innen auch zu einer sehr deutlichen und patientenrelevanten Reduktion der durch die Patient:innen wahrgenommenen Symptombelastung führt.

*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.*

## **Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

## **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S und PGI-C“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### 4.3.1.3.1.6 Morbidität – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)

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

Tabelle 4-41: Operationalisierung des Endpunkts „Morbidität – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“

Studie	Operationalisierung
BEYOND	<p>Der 6-Minuten-Gehtest (6MWT) ist ein standardisierter und etablierter Test zur Bestimmung körperlicher Belastbarkeit. Die Erhebung des 6MWT erfolgte zweimalig im Screening, d. h. in den 4 Wochen vor der ersten Dosis, zum Zeitpunkt jeder vierten Behandlungsvisite während der doppelblinden Behandlungsphase und zusätzlich zu Woche 24 und 48 (sofern eine Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der doppelblinden Behandlungsphase stattfand) sowie 9 Wochen nach der letzten Dosis.</p> <p>Die Auswertung des 6MWT erfolgte mittels ANCOVA basierend auf der Änderung bei der Visite zu Woche 48 gegenüber der letzten Messung vor Behandlungsbeginn (= Baseline), wobei eine Adjustierung für den Baseline-Wert sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren vorgenommen wurde. Zur Quantifizierung des Gruppenunterschieds wird die adjustierte Mittelwertdifferenz sowie der entsprechende Schätzer von Hedges' g betrachtet.</p> <p>Ergänzend wird die Änderung des 6MWT zu Woche 24 dargestellt.</p> <p>Alle Analysen zu diesem Endpunkt basieren auf der ITT-Population, wobei nur Patient:innen mit hinreichenden Daten für die Auswertung berücksichtigt wurden, d. h. Patient:innen mit Messwerten zu den erforderlichen Zeitpunkten.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>○ <b>Hauptanalyse:</b> Mittlere Änderung des 6MWT im Vergleich zu Baseline zu Woche 48 (ergänzend zu Woche 24)</li> </ul>
6MWT: 6-Minuten-Gehtest; ANCOVA: <i>analysis of covariance</i>	

*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 4-42: Bewertung des Verzerrungspotenzials für den „Endpunkt Morbidität – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“ in der Studie BEYOND

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

*Begründen Sie für jede Studie die abschließende Einschätzung.*

Der Endpunkt „Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung oder sonstige Aspekte, die zu einer Verzerrung führen könnten, liegen nicht vor.

Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

*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.*

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

**Hauptanalyse: Mittlere Änderung des 6MWT**

Tabelle 4-43: Ergebnisse für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)“ – Änderung des 6MWT gegenüber Baseline aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel (SE) <sup>(4)</sup>	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel (SE) <sup>(4)</sup>	Mittlere Differenz [95 %-KI]; p-Wert <sup>(4)</sup>	Hedges' g [95 %-KI] <sup>(5)</sup>
<b>Änderung des 6MWT gegenüber Baseline</b>								
Mittlere Änderung zu Woche 24 (in Metern)	87	446,97 (92,97)	7,20 (7,017)	47	434,98 (78,33)	-8,96 (9,297)	16,15 [-5,73; 38,04]; 0,1466	0,25 [-0,09; 0,60]
Mittlere Änderung zu Woche 48 (in Metern)	86	446,97 (92,97)	8,82 (5,907)	41	434,98 (78,33)	-3,62 (8,141)	12,44 [-6,72; 31,59]; 0,2011	0,22 [-0,12; 0,57]
ITT-Population; Primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.								
<sup>(2)</sup> Patient:innen mit einem 6MWT-Wert zu Baseline und zu Woche 24 bzw. 48								
<sup>(3)</sup> Arithmetischer Mittelwert basierend auf dem letzten Wert vor Therapiebeginn (Angabe bezieht sich auf alle randomisierten Patient:innen)								
<sup>(4)</sup> ANCOVA basierend auf der Änderung gegenüber Baseline, adjustiert für den Baseline-6MWT-Wert, den Baseline-Hb-Wert und den Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(5)</sup> Hedges' g abgeleitet aus dem Modellschätzer der mittleren Differenz								
ANCOVA: <i>analysis of covariance</i> ; BSC: <i>Best-Supportive-Care</i> ; Hb: Hämoglobin; ITT: Intention-to-Treat; KI: Konfidenzintervall; SD: Standardabweichung; SE: Standardfehler								

Für die körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest zeigt sich in der Studie BEYOND in der Analyse der mittleren Änderung gegenüber Baseline zu Woche 48 (wie auch zu Woche 24) kein signifikanter Unterschied zwischen den Behandlungsarmen (mittlere Differenz = 12,44 m; 95 %-KI: [-6,72; 31,59];  $p = 0,2011$ ; siehe Hauptanalyse: Mittlere Änderung des 6MWT, Tabelle 4-43)

*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.*

### **Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

### **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### **4.3.1.3.1.7 Lebensqualität gemäß SF-36**

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

Tabelle 4-44: Operationalisierung des Endpunkts „Lebensqualität – SF-36“

<b>Studie</b>	<b>Operationalisierung</b>
BEYOND	Die Erhebung der gesundheitsbezogenen Lebensqualität mittels SF-36 erfolgte einmalig im Screening, d. h. in den 4 Wochen vor Behandlungsbeginn, bei jeder zweiten Behandlungsvisite während der doppelblinden Behandlungsphase (beginnend mit der dritten Dosis), zusätzlich zu Woche 13, 24, 37 und 48 (sofern eine Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der doppelblinden Behandlungsphase stattfand), sowie 9 Wochen nach der letzten Dosis.  Für die Auswertung wurden die Summenscores zum körperlichen Wohlbefinden ( <i>Physical Component Summary</i> ; PCS) und zum psychischen Wohlbefinden ( <i>Mental Component Summary</i> ; MCS) berechnet. Höhere Werte in Summenscores des SF-36 entsprechen einer höheren Lebensqualität.  Die Auswertung erfolgte jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten

Studie	Operationalisierung
	<p>Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber dem Ausgangswert. Als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung wurden für den PCS 9,4 Punkte und für den MCS 9,6 Punkte verwendet. Dies entspricht der vom G-BA geforderten Responseschwelle von 15 % der Skalenspannweite. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell unter Berücksichtigung des Baseline-Werts sowie ein stratifizierter Log-Rank-Test berechnet. Die Stratifizierung erfolgte basierend auf den Stratifizierungsfaktoren der Randomisierung.</p> <p>Als Zusatzanalysen werden im Rahmen von Responderanalysen die Anteile der Patient:innen mit einer bewertungsrelevanten Verbesserung sowie einer bewertungsrelevanten Verschlechterung zu Woche 48 (und ergänzend zu Woche 24) dargestellt. Patient:innen mit einem fehlenden Wert zu Woche 48 (bzw. Woche 24) wurden von diesen Analysen ausgeschlossen. Als Schätzer des Behandlungseffekts wird das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.</p> <p>Um die mittlere Änderung der patientenberichteten Scores auch im zeitlichen Verlauf abzubilden, werden die arithmetischen Mittelwerte pro Behandlungsarm und Visite graphisch durch Verlaufskurven dargestellt.</p> <p>Weiterhin wurde für die Analyse der mittleren Änderung ein MMRM auf Basis der Daten aus dem Beobachtungszeitraum bis Woche 48 berechnet. Im entsprechenden Modell wurden, zusätzlich zur Behandlungsgruppe, der jeweilige Baseline-Score, die Visite, die Interaktion Behandlungsgruppe*Visite, die Interaktion Baseline-Score*Visite sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren als feste Effekte verwendet. Der Achsenschnittpunkt (Intercept) und die Visite wurden zudem als zufällige Effekte definiert. Zur Quantifizierung des Behandlungseffekts wird der adjustierte Modellschätzer der Mittelwertdifferenz zu Woche 48 sowie die hieraus abgeleitete standardisierte Mittelwertdifferenz (SMD) in Form von Hedges' g dargestellt.</p> <p>Die Analysen basieren auf der ITT-Population, wobei nur Patient:innen mit hinreichenden Daten für die jeweilige Auswertung berücksichtigt wurden, d. h. Patient:innen mit ausgefüllten Fragebögen zu den für die Analyse erforderlichen Zeitpunkten. Für den SF-36 gilt ein Fragebogen dann als ausgefüllt, wenn mindestens die Hälfte aller Fragen beantwortet wurden, d. h. mindestens 18 Fragen der 36 Items.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß SF-36</li> <li>• <b>Zusatzanalysen:</b> <ul style="list-style-type: none"> <li>○ Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß SF-36 zu Woche 24 und 48</li> <li>○ Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß SF-36 bis zum primären Datenschnitt</li> </ul> </li> <li>• <b>Zusatzanalysen in Anhang 4G:</b> <ul style="list-style-type: none"> <li>○ Rücklaufquoten zum SF-36</li> <li>○ Mittlere Änderung der gesundheitsbezogenen Lebensqualität gemäß SF-36 (MMRM-Analyse)</li> </ul> </li> </ul>
	<p>ITT: Intention-to-treat; KI: Konfidenzintervall; MCS: psychisches Wohlbefinden (<i>mental component summary</i>); MMRM: <i>mixed-model repeated measure</i>; PCS: körperliches Wohlbefinden (<i>physical component summary</i>); SF-36: <i>Short Form (36) Health Survey</i>; SMD: standardisierte Mittelwertdifferenzen</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 4-45: Bewertung des Verzerrungspotenzials den Endpunkt „Lebensqualität – SF-36“ in der Studie BEYOND

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
BEYOND	niedrig	ja	ja	ja	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt SF-36 wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert (siehe Tabelle 4-46).

Tabelle 4-46: Beobachtungszeiten für den Endpunkt „Lebensqualität – SF-36“ in der Studie BEYOND

Studie BEYOND Beobachtungszeit in Wochen	Luspatercept + BSC <sup>(1)</sup> (N=96)	Placebo + BSC <sup>(1)</sup> (N=49)
Mittelwert (SD)	88,4 (32,13)	69,0 (29,27)
Median (Min; Max)	95,2 (0,1; 130,4)	59,1 (0,1; 120,1)
Primärer Datenschnitt: 14.09.2020; ITT-Population; <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. BSC: <i>Best-Supportive-Care</i> ; ITT: Intention-to-Treat; Max: Maximum; Min: Minimum; SD: Standardabweichung		

Bei Patient:innen, die die Therapie abbrechen, wurde der SF-36 weiterhin bis Woche 48 bzw. 9 Wochen nach der letzten Dosis erhoben.

Für den SF-36 liegen sowohl für Woche 24, als auch für Woche 48 hohe Rücklaufquoten vor (über 70 % in beiden Behandlungsarmen, vgl. Anhang 4-G).

In Bezug auf Patient:innen, die die Therapie abbrechen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit

potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

*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.*

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

**Hauptanalyse: Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß SF-36**

Tabelle 4-47: Ergebnisse für den Endpunkt „Lebensqualität – SF-36“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
<b>Zeit bis zur ersten bewertungsrelevanten Verbesserung</b>								
PCS	94	27 (28,7)	N.A. [N.A.; N.A.]	47	5 (10,6)	N.A. [N.A.; N.A.]	2,79 [1,06; 7,38]	<b>0,0457</b>
MCS	94	32 (34,0)	N.A. [28,1; N.A.]	47	9 (19,1)	N.A. [N.A.; N.A.]	1,44 [0,68; 3,08]	0,1056
<b>Zeit bis zur ersten bewertungsrelevanten Verschlechterung</b>								
PCS	94	16 (17,0)	N.A. [N.A.; N.A.]	47	10 (21,3)	N.A. [19,6; N.A.]	0,66 [0,29; 1,50]	0,3364
MCS	94	26 (27,7)	N.A. [N.A.; N.A.]	47	20 (42,6)	N.A. [9,8; N.A.]	0,59 [0,33; 1,07]	<b>0,0303</b>
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.								
<sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und mindestens einer Post-Baseline Bewertung im Studienverlauf								
<sup>(3)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verbesserung/Verschlechterung.								
<sup>(4)</sup> Cox-Regressionsmodell adjustiert für den Ausgangswert und stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(5)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; MCS: psychisches Wohlbefinden ( <i>mental component summary</i> ); N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; PCS: körperliches Wohlbefinden ( <i>physical component summary</i> ); SF-36: <i>Short Form (36) Health Survey</i>								

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

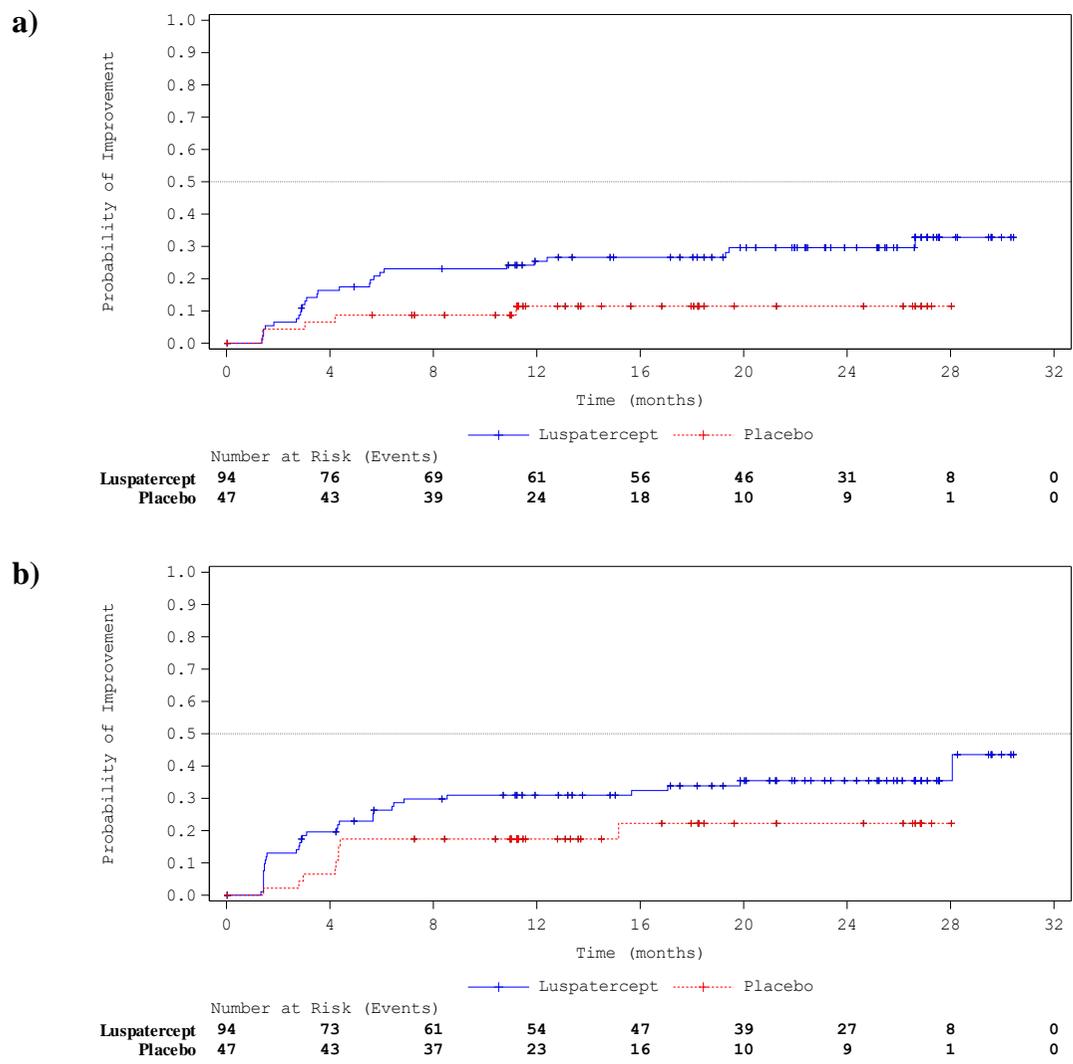


Abbildung 4-12: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung a) des SF-36 PCS und b) des SF-36 MCS

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

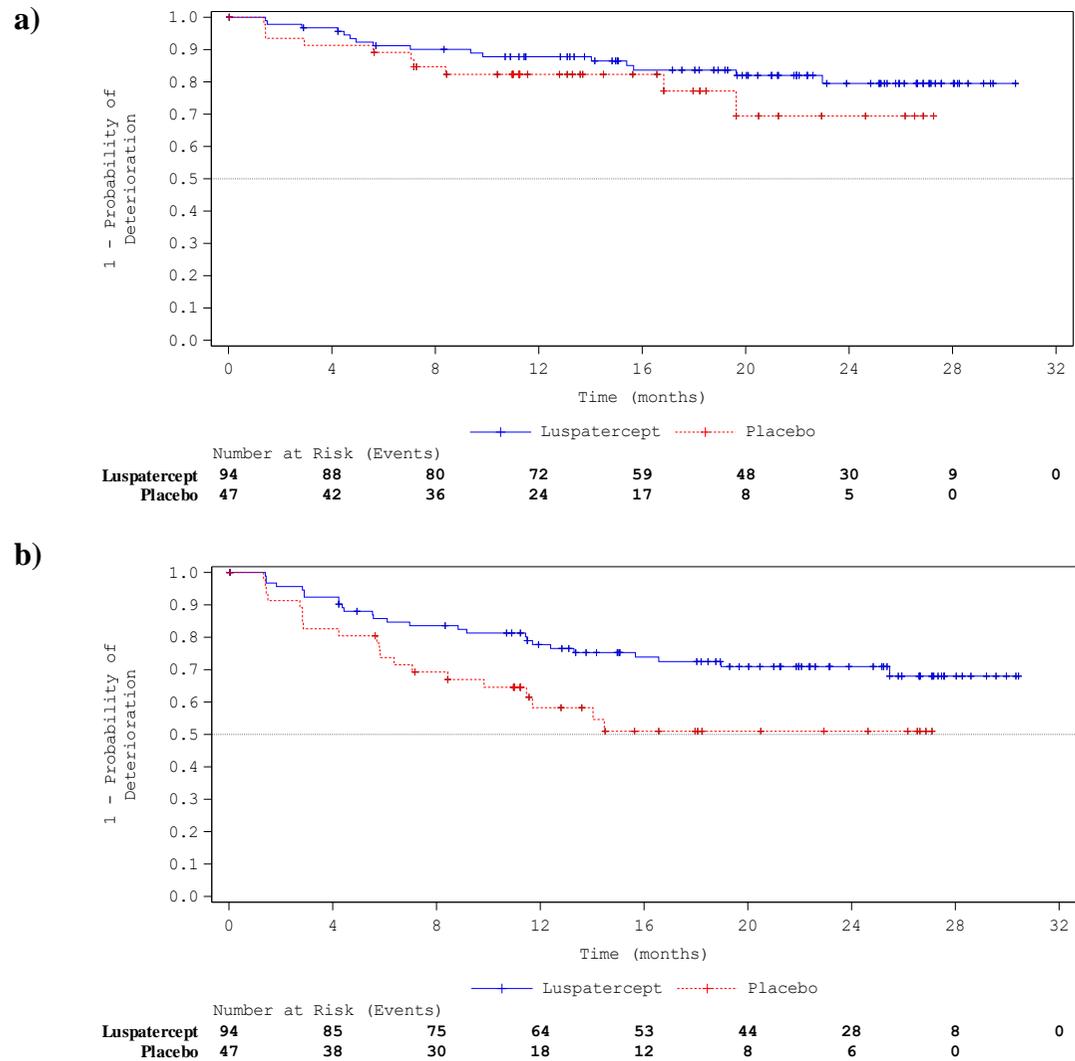


Abbildung 4-13: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung a) des SF-36 PCS und b) des SF-36 MCS

Für den Endpunkt SF-36 zeigten sich in der Studie BEYOND statistisch signifikante Effekte zugunsten von Luspatercept.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des SF-36 PCS (körperliches Wohlbefinden) ergab sich ein signifikanter Effekt mit einem HR von 2,79 (95 %-KI [1,06; 7,38];  $p = 0,0457$ ), d. h. die Chance auf eine bewertungsrelevante Verbesserung des körperlichen Wohlbefindens war unter Luspatercept mehr als doppelt so hoch wie unter Placebo. Wie in Abbildung 4-12a) ersichtlich, wurde der Median der Zeit bis zur ersten bewertungsrelevanten Verbesserung des PCS in beiden Behandlungsarmen zum Zeitpunkt des Datenschnitts noch nicht erreicht. In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verschlechterung des PCS zeigte sich kein signifikanter Unterschied.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des SF-36 MCS (psychisches Wohlbefinden) zeigte sich kein signifikanter Unterschied. In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verschlechterung des MCS ergab sich ein statistisch signifikanter Effekt mit einem HR von 0,59 (95 %-KI [0,33; 1,07];  $p = 0,0303$ ), d. h. das Risiko für eine bewertungsrelevante Verschlechterung des psychischen Wohlbefindens war unter Luspatercept 41 % geringer als unter Placebo. Wie in Abbildung 4-13b) ersichtlich, wurde der Median der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des MCS in beiden Behandlungsarmen zum Zeitpunkt des Datenschnitts noch nicht erreicht.

Weitere Zusatzanalysen sind im Folgenden dargestellt.

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

**Zusatzanalysen****Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß SF-36 zu Woche 24 und 48**

Tabelle 4-48: Ergebnisse für den Endpunkt „Lebensqualität – SF-36“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des SF-36 zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse)

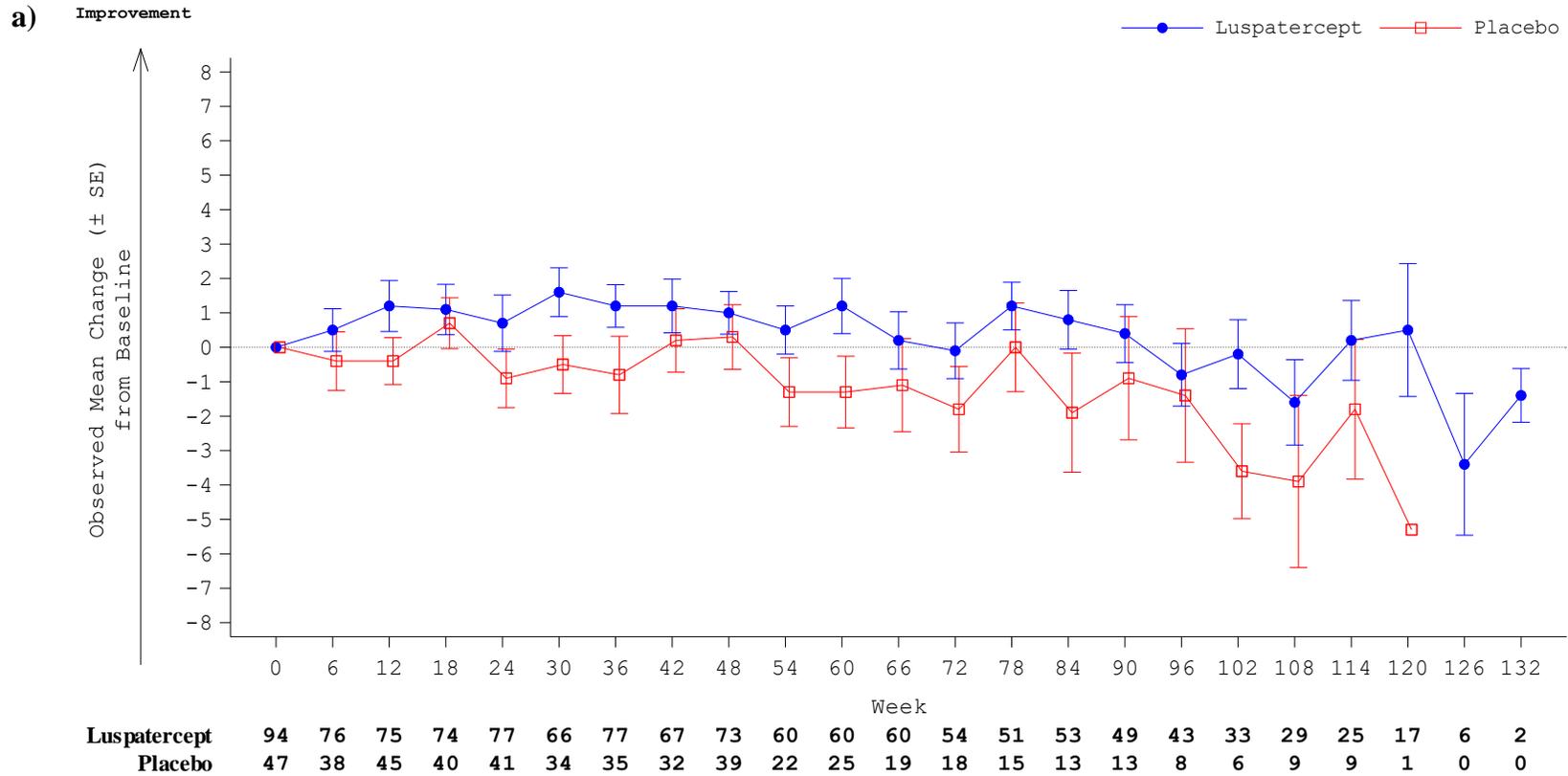
Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verbesserung des SF-36 PCS</b>							
Verbesserung zu Woche 24	77	10 (13,0)	41	0	0,13 [0,06; 0,21]	N.A.	N.A.
Verbesserung zu Woche 48	73	5 (6,8)	39	2 (5,1)	0,02 [-0,07; 0,10]	1,35 [0,24; 7,56]	1,31 [0,27; 6,26]; 0,7358
<b>Bewertungsrelevante Verbesserung des SF-36 MCS</b>							
Verbesserung zu Woche 24	77	10 (13,0)	41	2 (4,9)	0,08 [-0,01; 0,18]	3,15 [0,63; 15,62]	2,73 [0,63; 11,78]; 0,1783
Verbesserung zu Woche 48	73	11 (15,1)	39	1 (2,6)	0,12 [0,03; 0,22]	7,33 [0,88; 60,99]	5,93 [0,79; 44,22]; 0,0827
<b>Bewertungsrelevante Verschlechterung des SF-36 PCS</b>							
Verschlechterung zu Woche 24	77	6 (7,8)	41	2 (4,9)	0,03 [-0,06; 0,12]	1,58 [0,31; 7,97]	1,54 [0,34; 6,91]; 0,5763
Verschlechterung zu Woche 48	73	1 (1,4)	39	1 (2,6)	-0,01 [-0,07; 0,04]	0,54 [0,03; 8,78]	0,54 [0,03; 8,91]; 0,6639

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verschlechterung des SF-36 MCS</b>							
Verschlechterung zu Woche 24	77	7 (9,1)	41	4 (9,8)	0,00 [-0,11; 0,11]	0,99 [0,27; 3,63]	0,99 [0,31; 3,16]; 0,9874
Verschlechterung zu Woche 48	73	7 (9,6)	39	7 (17,9)	-0,08 [-0,22; 0,06]	0,49 [0,16; 1,50]	0,53 [0,20; 1,42]; 0,2078
ITT-Population; Primärer Datenschnitt vom 14.09.2020							
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und einer Bewertung zu Woche 24 bzw. 48 <sup>(3)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche <sup>(4)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet. BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; MCS: psychisches Wohlbefinden ( <i>mental component summary</i> ); N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; OR: Odds Ratio; PCS: körperliches Wohlbefinden ( <i>physical component summary</i> ); RD: Risiko Differenz; RR: Relatives Risiko; SF-36: <i>Short Form (36) Health Survey</i>							

In Bezug auf den Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung oder Verschlechterung der PCS und MCS zeigten sich weder zu Woche 24 noch zu Woche 48 statistisch signifikante Unterschiede zwischen den Behandlungsarmen (siehe Tabelle 4-48).

**Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß SF-36**



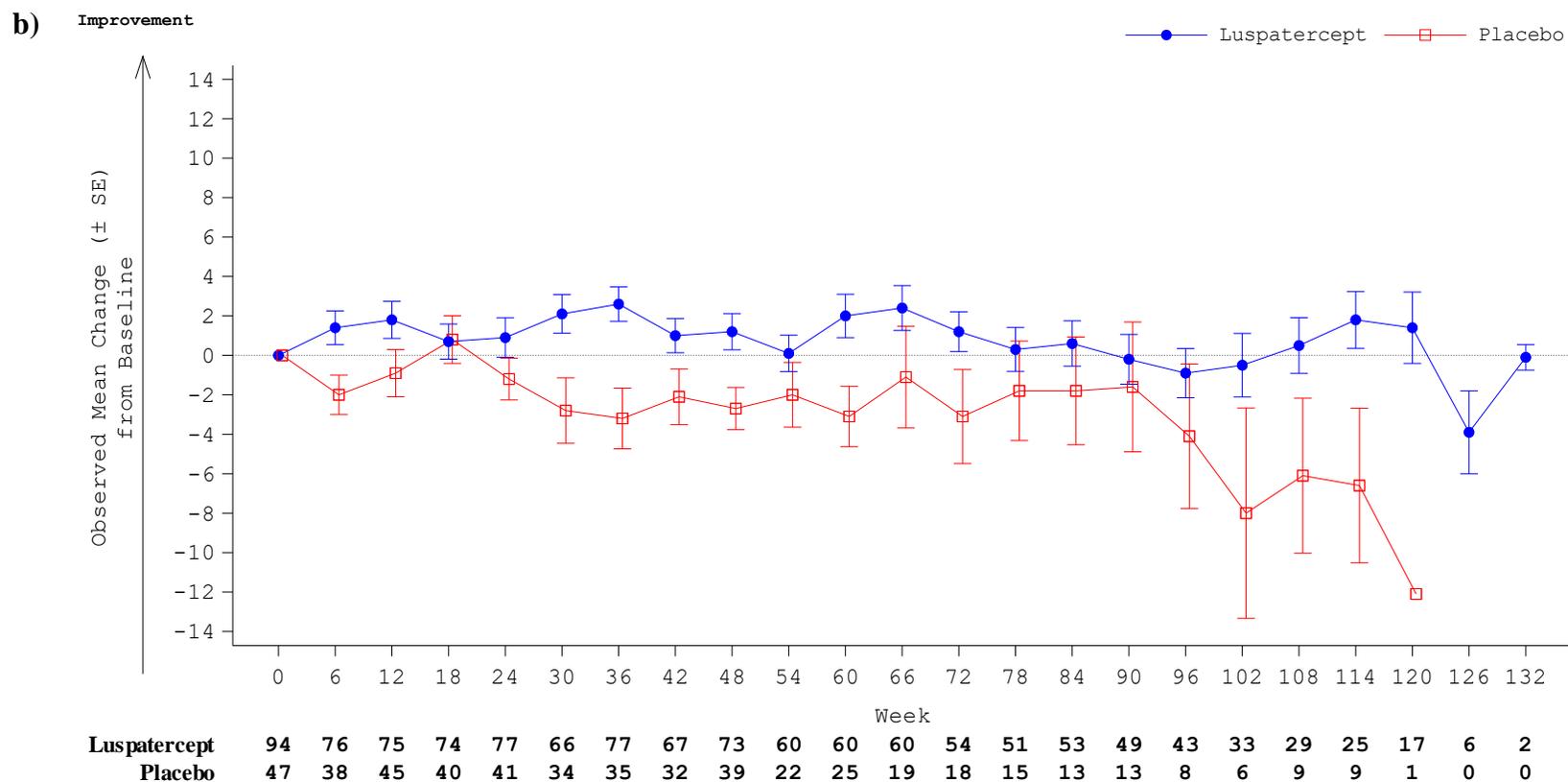


Abbildung 4-14: Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß a) SF-36 PCS und b) SF-36 MCS

Im Verlauf der mittleren Änderung des PCS und MCS deutet sich ein positiveres Abschneiden des Luspatercept-Arms gegenüber Placebo an (siehe Abbildung 4-14). Auch bei den MMRM-Analysen zur mittleren Änderung des PCS und MCS zeigte sich sowohl zu Woche 24 als auch zu Woche 48 jeweils ein tendenzieller Vorteil für Luspatercept (siehe Anhang 4-G).

**Zusammenfassung der Ergebnisse zum Endpunkt Lebensqualität gemessen anhand des generischen Fragebogens SF-36**

Die Analysen zum Endpunkt SF-36 zeigten statistisch signifikante Vorteile von Luspatercept. So ergab sich beim SF-36 PCS (körperliches Wohlbefinden) ein statistisch signifikanter Vorteil in Bezug auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung und beim SF-36 MCS (psychisches Wohlbefinden) ein statistisch signifikanter Vorteil in Bezug auf die Zeit bis zur ersten bewertungsrelevanten Verschlechterung.

Die Aufrechterhaltung oder Verbesserung der Lebensqualität ist im vorliegenden Anwendungsgebiet ein relevantes Therapieziel. Aus den Ergebnissen geht hervor, dass die anhaltende Linderung der Anämie entsprechend der Steigerung und langfristigen Stabilisierung der Hb-Werte durch Luspatercept auch zu einer patientenrelevanten Verbesserung der Lebensqualität führt.

*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.*

**Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

**Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Lebensqualität – SF-36“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

**4.3.1.3.1.8 Lebensqualität gemäß FACIT-F**

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

Tabelle 4-49: Operationalisierung des Endpunkts „Lebensqualität – FACIT-F“

Studie	Operationalisierung
BEYOND	<p>Die Erhebung der gesundheitsbezogenen Lebensqualität mittels des Fatigue-spezifischen Fragebogens FACIT-F erfolgte einmalig im Screening, d. h. in den 4 Wochen vor Behandlungsbeginn, bei jeder zweiten Behandlungsvisite während der doppelblinden Behandlungsphase (beginnend mit der dritten Dosis), zusätzlich zu Woche 13, 24, 37 und 48 (sofern eine Erhebung zu diesen Zeitpunkten nicht bereits im Rahmen der doppelblinden Behandlungsphase stattfand), sowie 9 Wochen nach der letzten Dosis.</p> <p>Ergebnisse aus der Fatigue-spezifischen Skala (FACIT-Fatigue Skala) und den Subskalen des FACT-G (physisches Wohlbefinden [PWB], soziales/familiäres Wohlbefinden [SWB], emotionales Wohlbefinden [EWB], funktionelles Wohlbefinden [FWB]) wurden im FACIT-F Gesamtscore zusammengefasst. Höhere Werte entsprechen einer höheren Lebensqualität.</p> <p>Die Auswertung erfolgte jeweils als Ereigniszeitanalyse basierend auf der Zeit bis zur ersten Verbesserung und der Zeit bis zur ersten Verschlechterung, definiert als die Zeit von der ersten Dosis bis zum ersten Auftreten einer bewertungsrelevanten Verbesserung bzw. Verschlechterung gegenüber dem Ausgangswert. Als Schwellenwert für eine bewertungsrelevante Verbesserung oder Verschlechterung wird die vom G-BA geforderte Responseschwelle von 15 % der Skalenspannweite herangezogen. Für den FACIT-F Gesamtscore entspricht dies einer Verbesserung/Verschlechterung um 24 Punkte. Für die Einzelnen Subskalen entspricht dies einer Verbesserung/Verschlechterung um 5 Punkte für PWB, SWB und FWB, 4 Punkte für EWB, 8 Punkte für die Fatigue-spezifische Skala und 17 Punkte für den FACT-G Gesamtscore. Patient:innen ohne eine entsprechende Verbesserung bzw. Verschlechterung wurden zum Zeitpunkt ihrer letzten verfügbaren Bewertung des jeweiligen Fragebogens zensiert. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell unter Berücksichtigung des Baseline-Werts sowie ein stratifizierter Log-Rank-Test berechnet. Die Stratifizierung erfolgte basierend auf den Stratifizierungsfaktoren der Randomisierung.</p> <p>Als Zusatzanalysen werden im Rahmen von Responderanalysen die Anteile der Patient:innen mit einer bewertungsrelevanten Verbesserung sowie einer bewertungsrelevanten Verschlechterung zu Woche 48 (und ergänzend zu Woche 24) dargestellt. Patient:innen mit einem fehlenden Wert zu Woche 48 (bzw. Woche 24) wurden von diesen Analysen ausgeschlossen. Als Schätzer des Behandlungseffekts wird das relative Risiko gemäß der Methode von Mantel-Haenszel mit Adjustierung für die Stratifizierungsfaktoren der Randomisierung dargestellt.</p> <p>Um die mittlere Änderung der patientenberichteten Scores auch im zeitlichen Verlauf abzubilden, werden die arithmetischen Mittelwerte pro Behandlungsarm und Visite graphisch durch Verlaufskurven dargestellt.</p> <p>Weiterhin wurde für die Analyse der mittleren Änderung ein MMRM auf Basis der Daten aus dem Beobachtungszeitraum bis Woche 48 berechnet. Im entsprechenden Modell wurden, zusätzlich zur Behandlungsgruppe, der jeweilige Baseline-Score, die Visite, die Interaktion Behandlungsgruppe*Visite, die Interaktion Baseline-Score*Visite sowie die für die Randomisierung verwendeten Stratifizierungsfaktoren als feste Effekte verwendet. Der Achsenschnittpunkt (Intercept) und die Visite wurden zudem als zufällige Effekte definiert. Zur Quantifizierung des Behandlungseffekts wird der adjustierte Modellschätzer der Mittelwertdifferenz zu Woche 48 (und ergänzend zu Woche 24) sowie die hieraus abgeleitete standardisierte Mittelwertdifferenz (SMD) in Form von Hedges' g darstellt.</p> <p>Die Analysen basieren auf der ITT-Population, wobei nur Patient:innen mit hinreichenden Daten für die jeweilige Auswertung berücksichtigt wurden, d. h. Patient:innen mit ausgefüllten Fragebögen zu den für die Analyse erforderlichen Zeitpunkten. Für den FACIT-F gilt ein Fragebogen dann als ausgefüllt, wenn mindestens 80 % aller Fragen beantwortet wurden, d. h. mindestens 32 der 40 Fragen.</p>

Studie	Operationalisierung
	<p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F Gesamtscore</li> <li>• <b>Zusatzanalysen:</b> <ul style="list-style-type: none"> <li>○ Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F Subskalen</li> <li>○ Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F Gesamtscore und Subskalen zu Woche 24 und 48</li> <li>○ Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß FACIT-F Gesamtscore bis zum primären Datenschnitt</li> </ul> </li> <li>• <b>Zusatzanalysen in Anhang 4G:</b> <ul style="list-style-type: none"> <li>○ Rücklaufquoten zum FACIT-F</li> <li>○ Mittlere Änderung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F Gesamtscore und Subskalen (MMRM-Analyse)</li> </ul> </li> </ul>
<p>ITT: Intention-to-treat; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy – Fatigue</i>; FACT-G: <i>Functional Assessment of Cancer Therapy – General</i>; KI: Konfidenzintervall; MMRM: <i>mixed-model repeated measure</i>; SMD: standardisierte Mittelwertdifferenzen</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 4-50: Bewertung des Verzerrungspotenzials für den Endpunkt „Lebensqualität – FACIT-F“ in der Studie BEYOND

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
BEYOND	niedrig	ja	ja	ja	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Der Endpunkt FACIT-F wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor.

Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert (siehe Tabelle 4-51).

Tabelle 4-51: Beobachtungszeiten für den Endpunkt „Lebensqualität – FACIT-F“ in der Studie BEYOND

<b>Studie BEYOND Beobachtungszeit in Wochen</b>	<b>Luspatercept + BSC <sup>(1)</sup> (N=96)</b>	<b>Placebo + BSC <sup>(1)</sup> (N=49)</b>
Mittelwert (SD)	88,4 (32,20)	69,0 (29,27)
Median (Min; Max)	95,2 (0,1; 130,4)	59,1 (0,1; 120,1)
Primärer Datenschnitt: 14.09.2020; ITT-Population; <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusions-therapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. BSC: <i>Best-Supportive-Care</i> ; ITT: Intention-to-Treat; Max: Maximum; Min: Minimum; SD: Standardabweichung		

Bei Patient:innen, die die Therapie abbrachen, wurde der FACIT-F weiterhin bis Woche 48 bzw. 9 Wochen nach der letzten Dosis erhoben.

Für den FACIT-F liegen sowohl für Woche 24, als auch für Woche 48 hohe Rücklaufquoten vor (über 70 % in beiden Behandlungsarmen, vgl. Anhang 4-G).

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensurierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

*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.*

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

**Hauptanalyse: Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F**

Tabelle 4-52: Ergebnisse für den Endpunkt „Lebensqualität – FACIT-F“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Hauptanalyse, Gesamtscore)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
<b>Zeit bis zur ersten bewertungsrelevanten Verbesserung</b>								
FACIT-F Gesamtscore	96	28 (29,2)	N.A. [N.A.; N.A.]	48	6 (12,5)	N.A. [N.A.; N.A.]	2,41 [0,98; 5,95]	<b>0,0340</b>
<b>Zeit bis zur ersten bewertungsrelevanten Verschlechterung</b>								
FACIT-F Gesamtscore	96	21 (21,9)	N.A. [N.A.; N.A.]	48	12 (25,0)	N.A. [N.A.; N.A.]	0,78 [0,38; 1,59]	0,5099
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie. <sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und mindestens einer Post-Baseline Bewertung im Studienverlauf <sup>(3)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verbesserung/Verschlechterung. <sup>(4)</sup> Cox-Regressionsmodell adjustiert für den Ausgangswert und stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche <sup>(5)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche BSC: <i>Best-Supportive-Care</i> ; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy – Fatigue</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

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

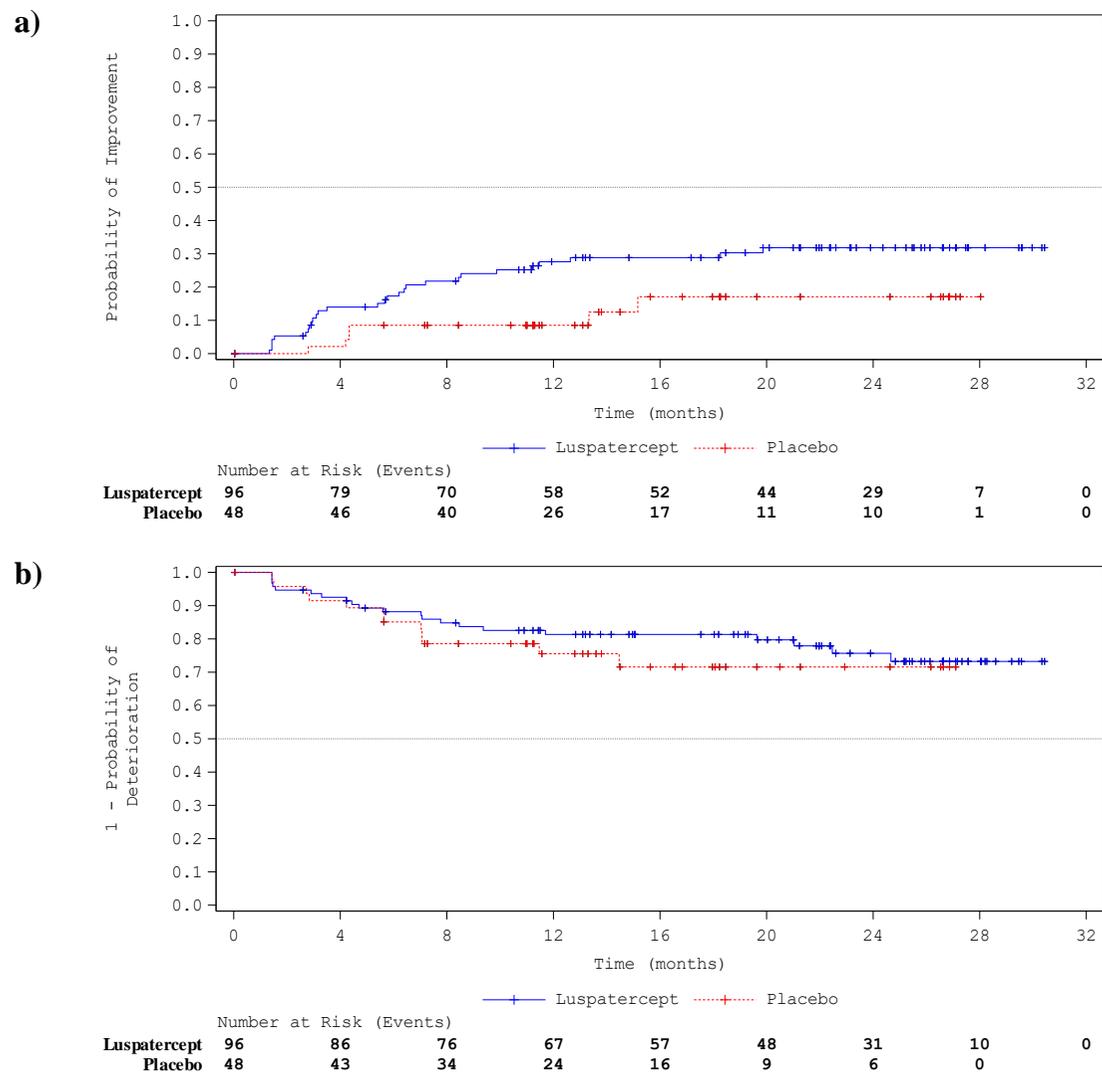


Abbildung 4-15: Kaplan-Meier Plot a) der Zeit bis zur ersten bewertungsrelevanten Verbesserung des FACIT-F Gesamtscores und b) der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des FACIT-F Gesamtscores

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

Für den Endpunkt FACIT-F zeigte sich in der Studie BEYOND ein statistisch signifikanter Effekt zugunsten von Luspatercept (siehe Tabelle 4-52).

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des FACIT-F Gesamtscores ergab sich ein statistisch signifikanter Effekt mit einem HR von 2,41 (95 %-KI [0,98; 5,95];  $p = 0,0340$ ), d. h. die Chance auf eine bewertungsrelevante Verbesserung der Lebensqualität gemäß FACIT-F war unter Luspatercept mehr als doppelt so hoch wie unter Placebo. Wie in Abbildung 4-15a) ersichtlich, wurde der Median der Zeit bis zur ersten bewertungsrelevanten Verbesserung in beiden Behandlungsarmen zum Zeitpunkt des Datenschnitts noch nicht erreicht.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verschlechterung zeigte sich kein signifikanter Unterschied.

Weitere Zusatzanalysen sind im Folgenden dargestellt.

### Zusatzanalysen

#### *Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F Subskalen*

Tabelle 4-53: Ergebnisse für den Endpunkt „Lebensqualität – FACIT-F“ – Zeit bis zur ersten bewertungsrelevanten Änderung gegenüber Baseline aus der Studie BEYOND (Zusatzanalyse, Subskalen)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
<b>Zeit bis zur ersten bewertungsrelevanten Verbesserung der FACIT-F Subskalen</b>								
<b>Physisches Wohlbefinden</b>	96	23 (24,0)	N.A. [N.A.; N.A.]	48	7 (14,6)	N.A. [N.A.; N.A.]	1,24 [0,52; 2,98]	0,2641
<b>Soziales/familiäres Wohlbefinden</b>	96	24 (25,0)	N.A. [N.A.; N.A.]	48	12 (25,0)	N.A. [18,0; N.A.]	0,74 [0,35; 1,56]	0,9291

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
<b>Emotionales Wohlbefinden</b>	96	28 (29,2)	N.A. [N.A.; N.A.]	48	10 (20,8)	N.A. [N.A.; N.A.]	1,51 [0,72; 3,14]	0,3326
<b>Funktionelles Wohlbefinden</b>	96	30 (31,3)	N.A. [N.A.; N.A.]	48	8 (16,7)	N.A. [N.A.; N.A.]	2,09 [0,95; 4,61]	<b>0,0475</b>
<b>FACT-G Gesamtscore</b>	96	21 (21,9)	N.A. [N.A.; N.A.]	48	5 (10,4)	N.A. [N.A.; N.A.]	1,56 [0,57; 4,29]	0,1411
<b>Fatigue-spezifische Skala</b>	96	38 (39,6)	N.A. [18,3; N.A.]	48	16 (33,3)	N.A. [21,8; N.A.]	1,21 [0,67; 2,20]	0,3777
<b>Zeit bis zur ersten bewertungsrelevanten Verschlechterung der FACIT-F Subskalen</b>								
<b>Physisches Wohlbefinden</b>	96	32 (33,3)	N.A. [25,7; N.A.]	48	16 (33,3)	N.A. [9,8; N.A.]	0,86 [0,47; 1,58]	0,6046
<b>Soziales/familiäres Wohlbefinden</b>	96	37 (38,5)	29,2 [21,9; N.A.]	48	13 (27,1)	N.A. [N.A.; N.A.]	1,34 [0,70; 2,54]	0,3662
<b>Emotionales Wohlbefinden</b>	96	29 (30,2)	N.A. [N.A.; N.A.]	48	14 (29,2)	N.A. [N.A.; N.A.]	0,84 [0,44; 1,59]	0,6072
<b>Funktionelles Wohlbefinden</b>	96	31 (32,3)	N.A. [27,1; N.A.]	48	16 (33,3)	26,2 [21,0; N.A.]	0,88 [0,48; 1,63]	0,5243
<b>FACT-G Gesamtscore</b>	96	21 (21,9)	N.A. [N.A.; N.A.]	48	12 (25,0)	N.A. [19,6; N.A.]	0,68 [0,33; 1,40]	0,3542
<b>Fatigue-spezifische Skala</b>	96	23 (24,0)	N.A. [N.A.; N.A.]	48	18 (37,5)	N.A. [8,5; N.A.]	0,55 [0,30; 1,03]	0,0585

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(3)</sup>	Hazard Ratio [95 %-KI] <sup>(4)</sup>	p-Wert <sup>(5)</sup>
ITT-Population; primärer Datenschnitt vom 14.09.2020 <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und mindestens einer Post-Baseline Bewertung im Studienverlauf <sup>(3)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur bewertungsrelevanten Verbesserung/Verschlechterung <sup>(4)</sup> Cox-Regressionsmodell adjustiert für den Ausgangswert und stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche <sup>(5)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche BSC: <i>Best-Supportive-Care</i> ; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy – Fatigue</i> ; FACT-G: <i>Functional Assessment of Cancer Therapy – General</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

Bei den Analysen der Zeit bis zur ersten bewertungsrelevanten Verbesserung bzw. Verschlechterung der FACIT-F Subskalen zeigte sich bzgl. der bewertungsrelevanten Verbesserung des funktionellen Wohlbefindens ein statistisch signifikanter Effekt zugunsten von Luspatercept (siehe Tabelle 4-53). Bei den anderen Subskalen und bei der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des funktionellen Wohlbefindens ergaben sich keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen.

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

**Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung der gesundheitsbezogenen Lebensqualität gemäß FACIT-F Gesamtscores und der Subskalen zu Woche 24 und 48**

Tabelle 4-54: Ergebnisse für den Endpunkt „Lebensqualität – FACIT-F“ – Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung bzw. Verschlechterung des FACIT-F Gesamtscores und Subskalen zu Woche 24 und 48 aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Bewertungsrelevante Verbesserung des FACIT-F Gesamtscores</b>							
Verbesserung zu Woche 24	78	8 (10,3)	41	2 (4,9)	0,05 [-0,04; 0,14]	2,30 [0,45; 11,73]	2,10 [0,46; 9,58]; 0,3371
Verbesserung zu Woche 48	74	9 (12,2)	40	1 (2,5)	0,10 [0,01; 0,19]	6,13 [0,71; 52,94]	4,99 [0,67; 36,94]; 0,1156
<b>Bewertungsrelevante Verschlechterung des FACIT-F Gesamtscores</b>							
Verschlechterung zu Woche 24	78	5 (6,4)	41	2 (4,9)	0,01 [-0,07; 0,09]	1,22 [0,22; 6,88]	1,19 [0,27; 5,20]; 0,8188
Verschlechterung zu Woche 48	74	4 (5,4)	40	3 (7,5)	-0,02 [-0,12; 0,07]	0,69 [0,14; 3,27]	0,71 [0,17; 3,01]; 0,6378
<b>Bewertungsrelevante Verbesserung der FACIT-F Subskalen</b>							
<b>Physisches Wohlbefinden</b>							
Verbesserung zu Woche 24	78	6 (7,7)	41	3 (7,3)	0,01 [-0,09; 0,10]	1,09 [0,25; 4,77]	1,08 [0,29; 3,95]; 0,9088
Verbesserung zu Woche 48	74	8 (10,8)	40	1 (2,5)	0,09 [0,00; 0,17]	5,21 [0,60; 45,08]	4,43 [0,59; 33,41]; 0,1484

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>Soziales/familiäres Wohlbefinden</b>							
Verbesserung zu Woche 24	78	12 (15,4)	41	6 (14,6)	0,01 [-0,13; 0,14]	1,05 [0,36; 3,03]	1,04 [0,42; 2,57]; 0,9278
Verbesserung zu Woche 48	74	8 (10,8)	40	4 (10,0)	0,01 [-0,11; 0,12]	1,11 [0,31; 4,05]	1,10 [0,36; 3,31]; 0,8695
<b>Emotionales Wohlbefinden</b>							
Verbesserung zu Woche 24	78	8 (10,3)	41	3 (7,3)	0,04 [-0,07; 0,14]	1,60 [0,38; 6,67]	1,51 [0,42; 5,38]; 0,5260
Verbesserung zu Woche 48	74	9 (12,2)	40	3 (7,5)	0,05 [-0,05; 0,15]	1,90 [0,44; 8,13]	1,66 [0,51; 5,43]; 0,3986
<b>Funktionelles Wohlbefinden</b>							
Verbesserung zu Woche 24	78	9 (11,5)	41	1 (2,4)	0,09 [0,00; 0,17]	5,38 [0,63; 46,29]	4,47 [0,62; 32,44]; 0,1383
Verbesserung zu Woche 48	74	5 (6,8)	40	1 (2,5)	0,04 [-0,03; 0,12]	2,87 [0,33; 24,98]	2,77 [0,33; 23,02]; 0,3445
<b>FACT-G Gesamtscore</b>							
Verbesserung zu Woche 24	78	5 (6,4)	41	2 (4,9)	0,01 [-0,07; 0,10]	1,33 [0,24; 7,38]	1,31 [0,26; 6,70]; 0,7480
Verbesserung zu Woche 48	74	5 (6,8)	40	1 (2,5)	0,04 [-0,03; 0,12]	3,08 [0,33; 28,87]	2,77 [0,35; 22,04]; 0,3345
<b>Fatigue-spezifische Skala</b>							
Verbesserung zu Woche 24	78	14 (17,9)	41	5 (12,2)	0,06 [-0,06; 0,19]	1,73 [0,55; 5,44]	1,54 [0,60; 3,94]; 0,3658

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
Verbesserung zu Woche 48	74	17 (23,0)	40	4 (10,0)	0,13 [0,01; 0,26]	3,07 [0,92; 10,29]	2,35 [0,88; 6,28]; 0,0879
<b>Bewertungsrelevante Verschlechterung der FACIT-F Subskalen</b>							
<b>Physisches Wohlbefinden</b>							
Verschlechterung zu Woche 24	78	10 (12,8)	41	5 (12,2)	0,01 [-0,12; 0,13]	1,07 [0,34; 3,34]	1,06 [0,39; 2,89]; 0,9115
Verschlechterung zu Woche 48	74	5 (6,8)	40	5 (12,5)	-0,06 [-0,18; 0,06]	0,50 [0,13; 1,86]	0,54 [0,16; 1,74]; 0,2999
<b>Soziales/familiäres Wohlbefinden</b>							
Verschlechterung zu Woche 24	78	12 (15,4)	41	5 (12,2)	0,03 [-0,10; 0,16]	1,29 [0,42; 3,95]	1,24 [0,46; 3,33]; 0,6642
Verschlechterung zu Woche 48	74	12 (16,2)	40	1 (2,5)	0,14 [0,04; 0,23]	7,82 [0,94; 64,80]	6,45 [0,87; 48,03]; 0,0686
<b>Emotionales Wohlbefinden</b>							
Verschlechterung zu Woche 24	78	4 (5,1)	41	5 (12,2)	-0,07 [-0,19; 0,04]	0,38 [0,10; 1,49]	0,41 [0,12; 1,41]; 0,1557
Verschlechterung zu Woche 48	74	7 (9,5)	40	6 (15,0)	-0,06 [-0,19; 0,07]	0,59 [0,18; 1,88]	0,63 [0,23; 1,73]; 0,3672
<b>Funktionelles Wohlbefinden</b>							
Verschlechterung zu Woche 24	78	7 (9,0)	41	6 (14,6)	-0,06 [-0,18; 0,06]	0,51 [0,15; 1,74]	0,57 [0,21; 1,56]; 0,2752
Verschlechterung zu Woche 48	74	8 (10,8)	40	3 (7,5)	0,03 [-0,08; 0,14]	1,47 [0,36; 5,94]	1,42 [0,40; 5,04]; 0,5903

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)		Placebo + BSC <sup>(1)</sup> (N=49)		Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		
	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	N <sup>(2)</sup>	Patient:innen mit Ereignis n (%)	RD <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	OR <sup>(3)</sup> [95 %-KI] <sup>(4)</sup>	RR <sup>(3)</sup> [95 %-KI]; p-Wert <sup>(4)</sup>
<b>FACT-G Gesamtscore</b>							
Verschlechterung zu Woche 24	78	6 (7,7)	41	2 (4,9)	0,02 [-0,07; 0,11]	1,47 [0,28; 7,66]	1,43 [0,31; 6,48]; 0,6439
Verschlechterung zu Woche 48	74	4 (5,4)	40	3 (7,5)	-0,02 [-0,12; 0,07]	0,69 [0,14; 3,27]	0,71 [0,17; 2,99]; 0,6354
<b>Fatigue-spezifische Skala</b>							
Verschlechterung zu Woche 24	78	9 (11,5)	41	4 (9,8)	0,02 [-0,10; 0,13]	1,19 [0,34; 4,11]	1,16 [0,39; 3,46]; 0,7837
Verschlechterung zu Woche 48	74	6 (8,1)	40	3 (7,5)	0,00 [-0,10; 0,11]	1,07 [0,25; 4,60]	1,06 [0,28; 3,98]; 0,9314
ITT-Population; Primärer Datenschnitt vom 14.09.2020							
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Patient:innen mit einer Bewertung zu Baseline und einer Bewertung zu Woche 24 bzw. 48 <sup>(3)</sup> RD, OR und RR mittels Mantel-Haenszel-Methode, adjustiert für Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche <sup>(4)</sup> KIs und p-Wert wurden mittels Normalapproximation berechnet. BSC: <i>Best-Supportive-Care</i> ; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy – Fatigue</i> ; FACT-G: <i>Functional Assessment of Cancer Therapy – General</i> ; KI: Konfidenzintervall; OR: Odds Ratio; RD: Risiko Differenz; RR: Relatives Risiko							

In Bezug auf den Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung oder Verschlechterung des FACIT-F Gesamtscores und der FACIT-F Subskalen zeigten sich weder zu Woche 24 noch zu Woche 48 statistisch signifikante Unterschiede zwischen den Behandlungsarmen (siehe Tabelle 4-54).

**Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß FACIT-F**

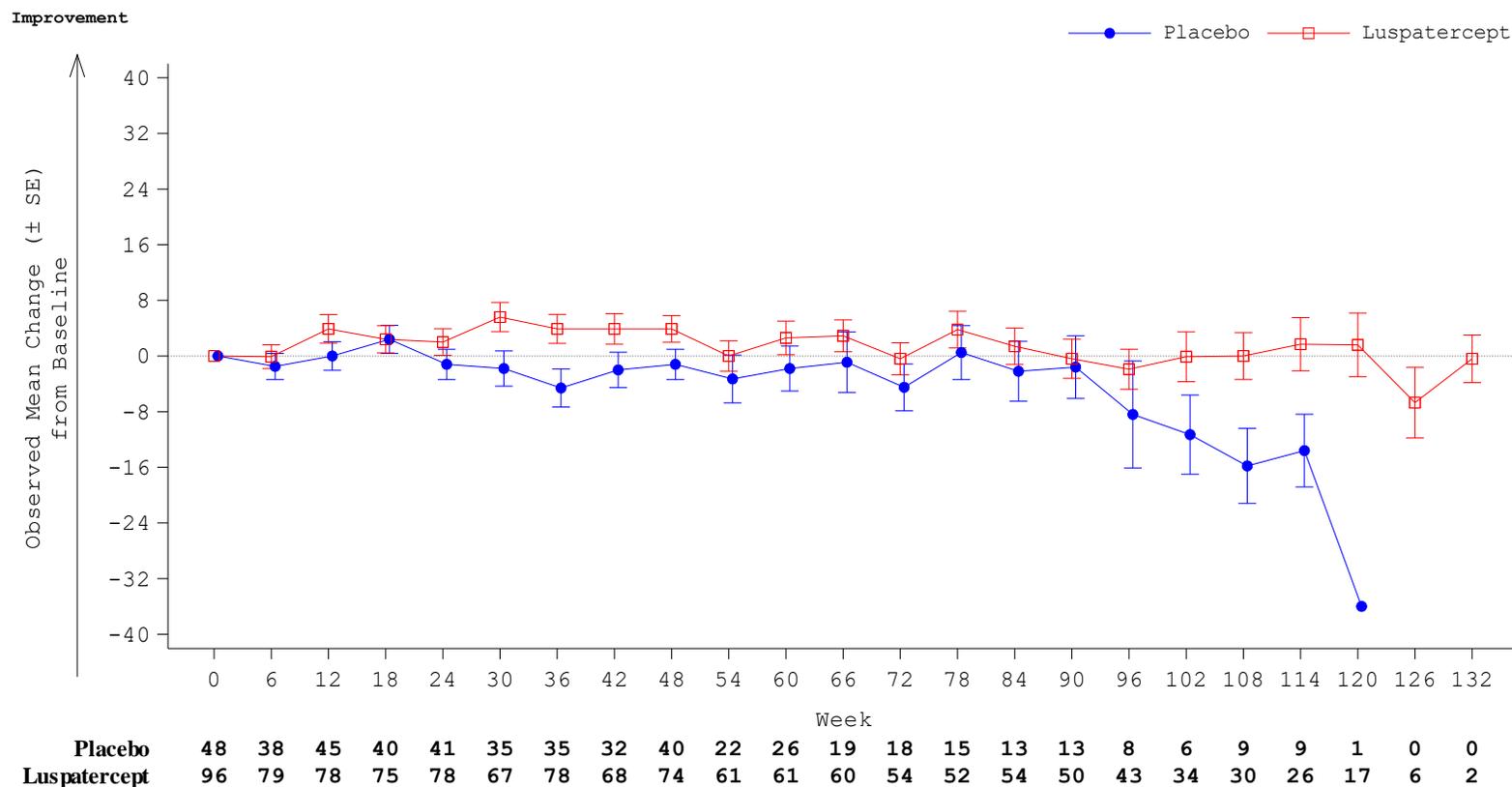


Abbildung 4-16: Line Plot zum Verlauf der mittleren Änderung der Lebensqualität gemäß des FACIT-F Gesamtscores

Im Verlauf der mittleren Änderung des FACIT-F Gesamtscores deutet sich ein positiveres Abschneiden des Luspatercept-Arms gegenüber Placebo an (siehe Abbildung 4-16). Auch bei den MMRM-Analysen zur mittleren Änderung des FACIT-F Gesamtscores und der Subskalen zeigten sich sowohl zu Woche 24 als auch zu Woche 48 tendenzielle Vorteile für Luspatercept (siehe Anhang 4-G).

**Zusammenfassung der Ergebnisse zum Endpunkt Lebensqualität gemessen anhand des Fatigue-spezifischen Fragebogens FACIT-F**

Beim FACIT-F Gesamtscore wie auch in der Subskala funktionelles Wohlbefinden zeigte sich jeweils ein statistisch signifikanter Vorteil von Luspatercept in Bezug auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung.

Die Aufrechterhaltung oder Verbesserung der Lebensqualität ist im vorliegenden Anwendungsgebiet ein relevantes Therapieziel. Aus den Ergebnissen geht hervor, dass die anhaltende Linderung der Anämie entsprechend der Steigerung und langfristigen Stabilisierung der Hb-Werte durch Luspatercept auch zu einer patientenrelevanten Verbesserung der Lebensqualität führt.

*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.*

**Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

**Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Lebensqualität – FACIT-F“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### 4.3.1.3.1.9 Sicherheit

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

Tabelle 4-55: Operationalisierung des Endpunkts „Sicherheit“

Studie	Operationalisierung
BEYOND	<p>Die Sicherheit wurde anhand des Auftretens von UE ausgewertet, die ab dem Tag der ersten Dosis Studienmedikation bis 63 Tage nach der letzten Dosis auftraten.</p> <p>UE wurden gemäß MedDRA (Version 23.0) kodiert und gemäß NCI CTCAE (Version 4.0) nach Schweregrad klassifiziert. UE, deren Schweregrad durch die CTCAE-Klassifizierung nicht definiert werden, wurden durch den oder die Prüfarzt:in anhand einer 5-Punkte-Skala bewertet (Grad 1: mild; Grad 2: moderat; Grad 3: schwer; Grad 4: lebensbedrohlich; Grad 5: tödlich).</p> <p>Als schwerwiegende UE (SUE) wurden solche Ereignisse berücksichtigt, die</p> <ul style="list-style-type: none"> <li>- tödlich verliefen,</li> <li>- lebensbedrohlich waren,</li> <li>- eine Hospitalisierung erforderlich machten oder zur Verlängerung eines bestehenden stationären Aufenthalts führten,</li> <li>- in bestehenden oder bedeutsamen Behinderungen/Invaliditäten resultierten,</li> <li>- angeborene Anomalien/Geburtsdefekte nach sich zogen oder</li> <li>- wichtige medizinische Ereignisse waren, die nicht unmittelbar tödlich oder lebensbedrohlich waren oder eine Hospitalisierung oder Behinderung verursachten, aber den Patienten bzw. die Patientin diesen Risiken aussetzten oder eine medizinische oder chirurgische Behandlung zur Prävention dieser Ereignisse nötig machten.</li> </ul> <p>Präspezifizierte UE von besonderem Interesse und weitere untersuchte UE werden anhand folgender Kategorien dargestellt:</p> <ul style="list-style-type: none"> <li>• UE von besonderem Interesse (AESI) <ul style="list-style-type: none"> <li>○ Malignitäten</li> <li>○ Prämaligne Erkrankungen</li> </ul> </li> <li>• Weitere untersuchte UE <ul style="list-style-type: none"> <li>○ Thromboembolische Ereignisse und Thrombophlebitis</li> <li>○ Knochenschmerzen und knochenschmerzähnliche Ereignisse</li> <li>○ Hypertonie</li> <li>○ Extramedulläre Hämatopoese</li> <li>○ Traumatische Frakturen</li> <li>○ Beinulcera</li> </ul> </li> </ul> <p>Im Rahmen der Auswertung werden im Dossier folgende UE-Kategorien betrachtet („UE-Hauptkategorien“):</p> <ul style="list-style-type: none"> <li>• Jegliche UE (nur ergänzend dargestellt)</li> <li>• SUE</li> <li>• Schwere UE (CTCAE-Grad <math>\geq 3</math>)</li> <li>• Zum Therapieabbruch führende UE</li> </ul> <p>Neben der Betrachtung der Gesamtraten erfolgt innerhalb der UE-Hauptkategorien eine Differenzierung nach SOC und PT. Zudem wurden im Rahmen von Zusatzanalysen die UE von besonderem Interesse und weitere untersuchte UE (jeweils insgesamt und separat für schwere und schwerwiegende Ereignisse dargestellt) und Hospitalisierungen ausgewertet.</p> <p>Als Hospitalisierungen wurden alle dokumentierten stationären Krankenhausaufenthalte während der Studie unabhängig von ihrer Ursache gezählt. Zusätzlich erfolgte eine deskriptive Auswertung der Gründe aller Hospitalisierungen sowie eine Auswertung der Hospitalisierungen</p>

Studie	Operationalisierung
	<p>ohne Berücksichtigungen von Hospitalisierungen aufgrund von UE.</p> <p>Die Auswertung erfolgte jeweils anhand einer Ereigniszeitanalyse auf Basis der Zeit von der ersten Dosis bis zum ersten UE innerhalb der betrachteten Kategorie. Zur Quantifizierung des Behandlungseffekts wurden das HR mit einem stratifizierten Cox-Regressionsmodell sowie ein stratifizierter Log-Rank-Test berechnet. Patient:innen ohne ein entsprechendes UE wurden 63 Tage nach der letzten Dosis bzw. zum Zeitpunkt des betrachteten Datenschnitts oder bei vorherigem Studienabbruch/Tod, je nachdem welches Ereignis früher eintrat, zensiert</p> <p>Auswertungen zu UE beziehen sich auf die Safety-Population, welche alle Patient:innen der ITT-Population umfasst, die mindestens eine Dosis Studienmedikation erhalten haben. Die Auswertungen zu Hospitalisierungen beziehen sich auf die ITT-Population.</p> <p><b>Darstellung der Ergebnisse:</b></p> <ul style="list-style-type: none"> <li>• <b>Hauptanalyse:</b> <ul style="list-style-type: none"> <li>○ Zeit bis zum ersten UE pro Hauptkategorie</li> <li>○ Zeit bis zum ersten UE auf Ebene der SOC und PT, bei denen sich ein signifikanter Unterschied zwischen den Studienarmen ergab (einschließlich Unterteilung hinsichtlich schwerwiegender und schwerer Ereignisse)</li> </ul> </li> <li>• <b>Zusatzanalysen:</b> <ul style="list-style-type: none"> <li>○ Zeit bis zur ersten Hospitalisierung (unabhängig von der Ursache und ohne Berücksichtigung von Hospitalisierungen aufgrund von UE)</li> <li>○ Deskriptive Zusammenfassung der Gründe für eine Hospitalisierung</li> </ul> </li> <li>• <b>Zusatzanalysen in Anhang 4 G:</b> <ul style="list-style-type: none"> <li>○ Zeit bis zum ersten UE auf Ebene der SOC und PT (separat für jegliche UE, SUE und schwere UE)</li> <li>○ Zeit bis zum ersten UE pro Kategorie der UE von besonderem Interesse und weiteren untersuchten UE (einschließlich Unterteilung hinsichtlich schwerwiegender und schwerer Ereignisse)</li> <li>○ Zum Therapieabbruch führende UE auf Ebene der SOC und PT</li> </ul> </li> </ul>
	<p>BSC: <i>Best-Supportive-Care</i>; CTCAE: <i>Common Terminology Criteria for Adverse Events</i>; HR: Hazard Ratio; ITT: Intention-to-Treat; KI: Konfidenzintervall; MedDRA: <i>Medical Dictionary for Regulatory Activities</i>; PT: bevorzugter Begriff (<i>preferred term</i>); SOC: Systemorganklasse (<i>system organ class</i>); SUE: schwerwiegendes UE; 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 4-56: Bewertung des Verzerrungspotenzials für den Endpunkt „Sicherheit“ in der Studie BEYOND

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
BEYOND	niedrig	ja	ja	ja	nein	hoch

*Begründen Sie für jede Studie die abschließende Einschätzung.*

Der Endpunkt wurde verblindet erhoben. In die Analyse der UE gingen solche Patient:innen ein, die mindestens eine Dosis Studienmedikation erhalten haben (Safety-Population). In der vorliegenden Studie ergaben sich jedoch zwischen der Safety-Population und der ITT-Population keine Unterschiede. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Entsprechend der Nachbeobachtung für unerwünschte Ereignisse von 63 Tagen nach der letzten Dosis, ergaben sich aus dem niedrigeren Anteil der Therapieabbrucher:innen und der längeren Therapiedauer unter Luspatercept im Vergleich zu Placebo unterschiedliche Beobachtungsdauern (vgl. Angaben zur Therapiedauer in Abschnitt 4.3.1.2.1). Diese Unterschiede wurden durch Ereigniszeitanalysen adressiert.

In Bezug auf Patient:innen, die die Therapie abbrachen, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

*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.*

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

**Hauptanalyse: Zeit bis zum ersten UE pro Hauptkategorie (UE-Gesamtraten und auf Ebene der SOC und PT)**

Tabelle 4-57: Ergebnisse für den Endpunkt „Sicherheit“ – Zeit bis zum ersten UE pro Hauptkategorie aus der Studie BEYOND (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
<b>Zeit bis zum ersten UE pro Hauptkategorie</b>								
Jegliche UE	96	96 (100,0)	0,10 [0,07; 0,13]	49	48 (98,0)	0,76 [0,46; 0,89]	1,47 [1,03; 2,12]	<b>0,0433</b>
SUE	96	11 (11,5)	N.A.	49	12 (24,5)	N.A. [18,00; N.A.]	0,29 [0,12; 0,69]	<b>0,0032</b>
Schwere UE (CTCAE ≥ 3)	96	27 (28,1)	N.A.	49	12 (24,5)	N.A. [16,62; N.A.]	1,07 [0,54; 2,14]	0,8419
Zum Therapieabbruch führende UE	96	3 (3,1)	N.A.	49	4 (8,2)	N.A.	0,29 [0,06; 1,34]	0,0923
Safety-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.								
<sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zum ersten UE								
<sup>(3)</sup> Cox-Regressionsmodell stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(4)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; SUE: schwerwiegendes UE; UE: unerwünschtes Ereignis								

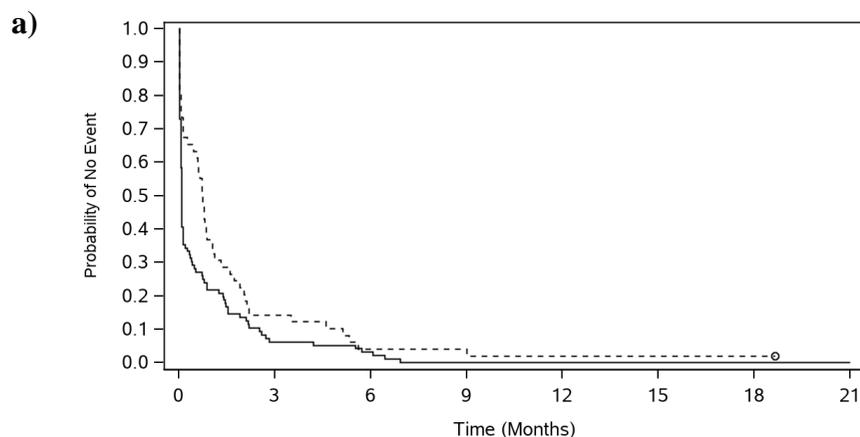
Bei den UE-Hauptkategorien zeigte sich unter Luspatercept gegenüber Placebo zum einen ein signifikanter Nachteil in Bezug auf jegliche UE und zum anderen ein signifikanter Vorteil in Bezug auf SUE. Bei den Analysen zu schweren UE und zum Therapieabbruch führende UE zeigten sich keine statistisch signifikanten Unterschiede (siehe Tabelle 4-57).

Die Auswertung zu jeglichen UE wird hier ergänzend betrachtet, gilt aber nicht per se als patientenrelevant (da hier z. B. auch milde UE, die ausschließlich auf Laborbefunde zurückgehen, berücksichtigt sind). Bei fast allen Patient:innen (100 % im Luspatercept-Arm, 98 % im

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

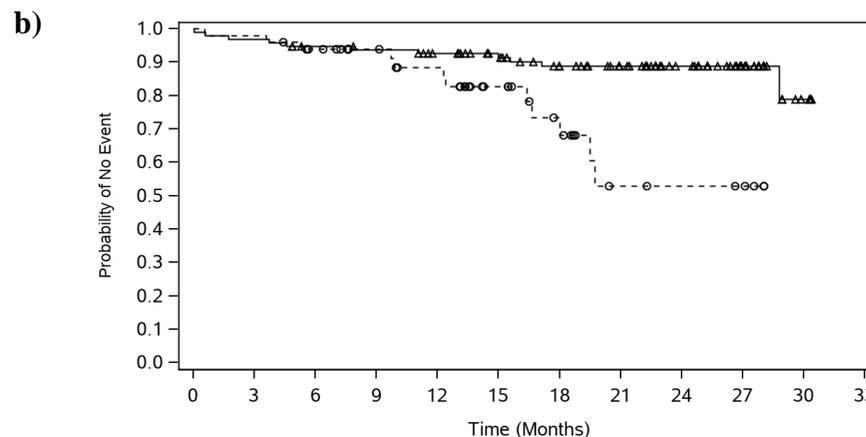
Placebo-Arm) trat mindestens ein UE auf. Basierend auf der Ereigniszeitanalyse war das Risiko für jegliche UE unter Luspatercept gegenüber Placebo um 47 % erhöht (HR [95 %-KI] = 1,47 [1,03; 2,12]; p = 0,0433). Wie in Abbildung 4-17a) ersichtlich, betrug die mediane Zeit bis zum ersten UE unter Luspatercept 0,10 Monate (95 %-KI [0,07; 0,13]) und unter Placebo 0,76 Monate (95 %-KI [0,46; 0,89]).

SUE traten bei 11,5 % der Patient:innen im Luspatercept-Arm und 24,5 % der Patient:innen im Placebo-Arm auf. Basierend auf der Ereigniszeitanalyse war das Risiko für SUE unter Luspatercept gegenüber Placebo um 71 % reduziert (HR [95 %-KI] = 0,29 [0,12; 0,69]; p = 0,0032). Zum Zeitpunkt des Datenschnitts war der Median der Zeit bis zum ersten SUE weder im Luspatercept-Arm noch im Placebo-Arm erreicht (siehe Abbildung 4-17b).



Number of Subjects at Risk							
Luspatercept							
96	6	3	0	0	0	0	0
Placebo							
49	7	2	2	1	1	1	0

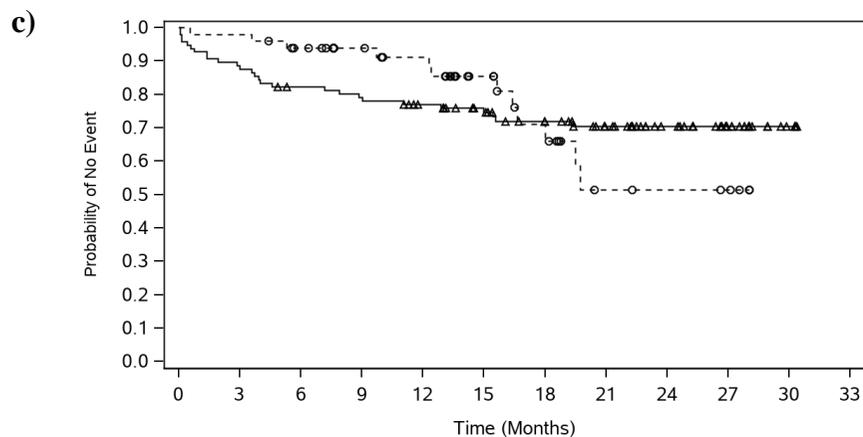
—▲— Luspatercept (events: 96/96), median and 95% CI: 0.10 (0.07, 0.13)  
 --○-- Placebo (events: 48/49), median and 95% CI: 0.76 (0.46, 0.89)  
 Luspatercept vs. Placebo - hazard ratio: 1.47 (1.03, 2.12), pvalue: 0.0433



Number of Subjects at Risk											
Luspatercept											
96	93	88	86	81	73	62	54	39	23	4	0
Placebo											
49	48	43	36	31	21	14	6	5	4	0	0

—▲— Luspatercept (events: 11/96), median and 95% CI: N.A.  
 --○-- Placebo (events: 12/49), median and 95% CI: N.A. (18.00, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 0.29 (0.12, 0.69), pvalue: 0.0032

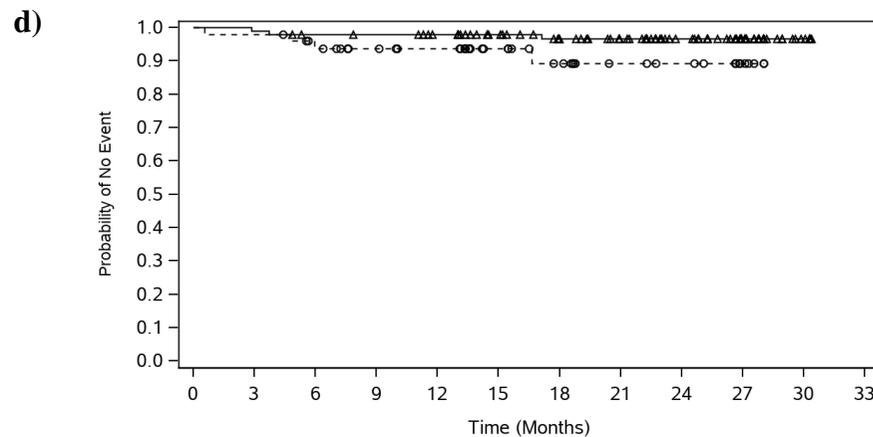
Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen



Number of Subjects at Risk

Luspatercept	96	84	77	74	68	60	51	42	29	17	4	0
Placebo	49	48	43	36	32	21	14	6	5	4	0	0

—▲— Luspatercept (events: 27/96), median and 95% CI: N.A.  
 --○-- Placebo (events: 12/49), median and 95% CI: N.A. (16.62, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 1.07 (0.54, 2.14), pvalue: 0.8419



Number of Subjects at Risk

Luspatercept	96	95	92	91	87	79	68	60	44	27	5	0
Placebo	49	48	43	38	35	25	19	13	11	5	0	0

—▲— Luspatercept (events: 3/96), median and 95% CI: N.A.  
 --○-- Placebo (events: 4/49), median and 95% CI: N.A.  
 Luspatercept vs. Placebo - hazard ratio: 0.29 (0.06, 1.34), pvalue: 0.0923

Abbildung 4-17: Kaplan-Meier Plot der Zeit bis zum ersten UE für a) jegliche UE, b) SUE c) schwere UE (CTCAE ≥ 3) und d) zum Therapieabbruch führende UE

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

Tabelle 4-58: Ergebnisse für den Endpunkt „Sicherheit“ – Zeit bis zum ersten UE auf Ebene der SOC und PT aus der Studie BEYOND, die bei mindestens 10 % bzw. 5 % der Patient:innen pro Studienarm auftraten und bei denen sich mindestens auf Ebene der PT ein signifikanter Unterschied zwischen den Studienarmen ergab (Hauptanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
<b>Jegliche UE, die bei mindestens 10 % der Patient:innen in einem Studienarm auftraten</b>								
SOC Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	96	76 (79,2)	0,72 [0,13; 2,07]	49	28 (57,1)	8,51 [1,74; N.A.]	1,85 [1,19; 2,86]	<b>0,0065</b>
PT Knochenschmerzen	96	35 (36,5)	N.A.	49	3 (6,1)	N.A.	7,11 [2,18; 23,15]	<b>0,0001</b>
SOC Infektionen und parasitäre Erkrankungen	96	71 (74,0)	7,62 [5,52; 10,87]	49	33 (67,3)	7,36 [3,25; 10,32]	0,96 [0,63; 1,46]	0,8469
PT Tonsillitis	96	2 (2,1)	N.A.	49	6 (12,2)	N.A.	0,13 [0,03; 0,64]	<b>0,0033</b>
SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums	96	40 (41,7)	N.A. [16,76; N.A.]	49	15 (30,6)	N.A.	1,25 [0,69; 2,27]	0,4718
PT Husten	96	17 (17,7)	N.A.	49	1 (2,0)	N.A.	7,24 [0,96; 54,68]	<b>0,0248</b>
SOC Gefäßerkrankungen	96	37 (38,5)	N.A. [25,07; N.A.]	49	10 (20,4)	N.A.	1,97 [0,98; 3,97]	0,0530
PT Hypertonie	96	19 (19,8)	N.A.	49	1 (2,0)	N.A.	9,32 [1,25; 69,66]	<b>0,0079</b>
SOC Erkrankungen der Haut und des Unterhautgewebes	96	26 (27,1)	N.A.	49	5 (10,2)	N.A.	2,59 [0,99; 6,77]	<b>0,0434</b>

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
<b>SUE, die bei mindestens 5 % der Patient:innen in einem Studienarm auftraten</b>								
SOC Infektionen und parasitäre Erkrankungen	96	2 (2,1)	N.A.	49	6 (12,2)	N.A. [19,48; N.A.]	0,11 [0,02; 0,57]	<b>0,0017</b>
<b>Schwere UE (CTCAE ≥ 3), die bei mindestens 5 % der Patient:innen in einem Studienarm auftraten</b>								
SOC Infektionen und parasitäre Erkrankungen	96	1 (1,0)	N.A.	49	6 (12,2)	N.A. [19,48; N.A.]	0,05 [0,01; 0,42]	<b>0,0002</b>
Safety-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie.								
<sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zum ersten UE								
<sup>(3)</sup> Cox-Regressionsmodell stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(4)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

Auf Ebene der SOC und PT zeigten sich im Vergleich von Luspatercept zu Placebo sowohl Vor- als auch Nachteile (siehe Tabelle 4-58).

Effekte zugunsten von Luspatercept liegen jeweils bei schweren und schwerwiegenden UE der SOC „Infektionen und parasitäre Erkrankungen“ und bei UE des PT „Tonsillitis“ vor.

Effekte zuungunsten von Luspatercept ergaben sich für UE der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“, dem PT „Knochenschmerzen“, dem PT „Hypertonie“, der SOC „Erkrankungen der Haut und des Unterhautgewebes“ und dem PT „Husten“. Knochenschmerzen und Hypertonie sind als Nebenwirkungen von Luspatercept bekannt. Knochenschmerzen traten bei 36,5 % der Patient:innen im Luspatercept-Arm und bei 6,1 % der Patient:innen im Placebo-Arm auf. Der Ereigniszeitanalyse entsprechend war das Risiko für Knochenschmerzen unter Luspatercept ca. 7-mal höher als unter Placebo (HR [95 %-KI] = 7,11 [2,18; 23,15]; p = 0,0001). Eine Hypertonie wiesen 19,8 % der Patient:innen im Luspatercept-Arm und 2,0 % der Patient:innen im Placebo-Arm auf. Der Ereigniszeitanalyse

---

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

entsprechend war das Risiko für Hypertonie unter Luspatercept ca. 9-mal höher als unter Placebo (HR [95 %-KI] = 9,32 [1,25; 69,66];  $p = 0,0079$ ).

Kaplan-Meier Plots der Zeit bis zum ersten UE pro Hauptkategorie auf Ebene der SOC und PT finden sich in Anhang 4-G. Weitere Zusatzanalysen sind im Folgenden dargestellt.

### **Zusatzanalysen**

#### *UE von besonderem Interesse sowie weitere untersuchte UE*

Als UE von besonderem Interesse wurden in der Studie BEYOND „Malignitäten“ und „Prämaligne Erkrankungen“ festgelegt. Als weitere UE berücksichtigt wurden analog zum CSR „Thromboembolische Ereignisse und Thrombophlebitis“, „Knochenschmerzen und knochen-schmerzähnliche Ereignisse“, „Hypertonie“, „Extramedulläre Hämatopoese“ und „Traumatische Frakturen“ sowie darüber hinaus „Beinulcera“ als zusätzliche patientenrelevante UE-Kategorie (vgl. Abschnitt 4.2.5.4). Die Auswertungen hierzu sind vollständig in Anhang 4-G dargestellt.

Im Hinblick auf „Prämaligne Erkrankungen“, „Thromboembolische Ereignisse und Thrombophlebitis“, „Extramedulläre Hämatopoese“, „Traumatische Frakturen“ sowie „Beinulcera“ ergaben sich keine Unterschiede zwischen den Studienarmen.

Bei der Auswertung zu „Malignitäten“ ergab sich in der Ereigniszeitanalyse gemäß Log-Rank-Test ein signifikanter Effekt zugunsten von Luspatercept ( $p = 0,0109$ ), das Hazard Ratio war nicht berechenbar. Insgesamt wiesen nur 2 Patient:innen eine Malignität auf, beide davon im Placebo-Arm (4,1 %). Eine der Malignitäten trat bereits 17 Tage nach Therapiebeginn auf, weshalb die Aussagekraft der vorliegenden Analyse zu Malignitäten fraglich ist.

Die Auswertung zur Hypertonie umfasst lediglich den PT „Hypertonie“ und entspricht somit der Auswertung der Hypertonie im vorhergehenden Abschnitt (siehe Tabelle 4-58).

Bei der Auswertung zu Knochenschmerzen und knochen-schmerzähnlichen Ereignissen, die auf einer ad hoc Zusammenstellung relevanter PTs beruht („Rückenschmerzen“, „Knochenschmerzen“, „Schmerzen des Muskel- und Skelettsystems“, „Nackenschmerzen“, „Schmerz in einer Extremität“, „Kieferschmerzen“ und „Wirbelsäulenschmerz“), zeigte sich ein signifikanter Effekt zuungunsten von Luspatercept (HR [95 %-KI] = 3,16 [1,80; 5,56];  $p < 0,0001$ ). Die Ergebnisse sind somit konsistent zur Auswertung des PT „Knochenschmerzen“ im vorhergehenden Abschnitt.

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

**Zeit bis zur ersten Hospitalisierung**

Tabelle 4-59: Ergebnisse für den Endpunkt „Sicherheit“– Zeit bis zur ersten Hospitalisierung aus der Studie BEYOND (Zusatzanalyse)

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Wochen) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Wochen) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
<b>Zeit bis zur ersten Hospitalisierung (unabhängig von der Ursache)</b>								
Jegliche Hospitalisierungen	96	16 (16,7)	N.A. [125,29; N.A.]	49	12 (24,5)	N.A. [84,71; N.A.]	0,50 [0,23; 1,08]	0,0702
<b>Zeit bis zur ersten Hospitalisierung, ohne Berücksichtigung von Hospitalisierungen aufgrund von UE</b>								
Nicht durch UE bedingte Hospitalisierungen	96	6 (6,3)	N.A. [N.A.; N.A.]	49	2 (4,1)	N.A. [N.A.; N.A.]	1,35 [0,27; 6,78]	0,7168
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.								
<sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten Hospitalisierung								
<sup>(3)</sup> Cox-Regressionsmodell stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
<sup>(4)</sup> Log-Rank-Test stratifiziert nach Baseline-Hb-Wert und Baseline-NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche								
BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; UE: unerwünschtes Ereignis								

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

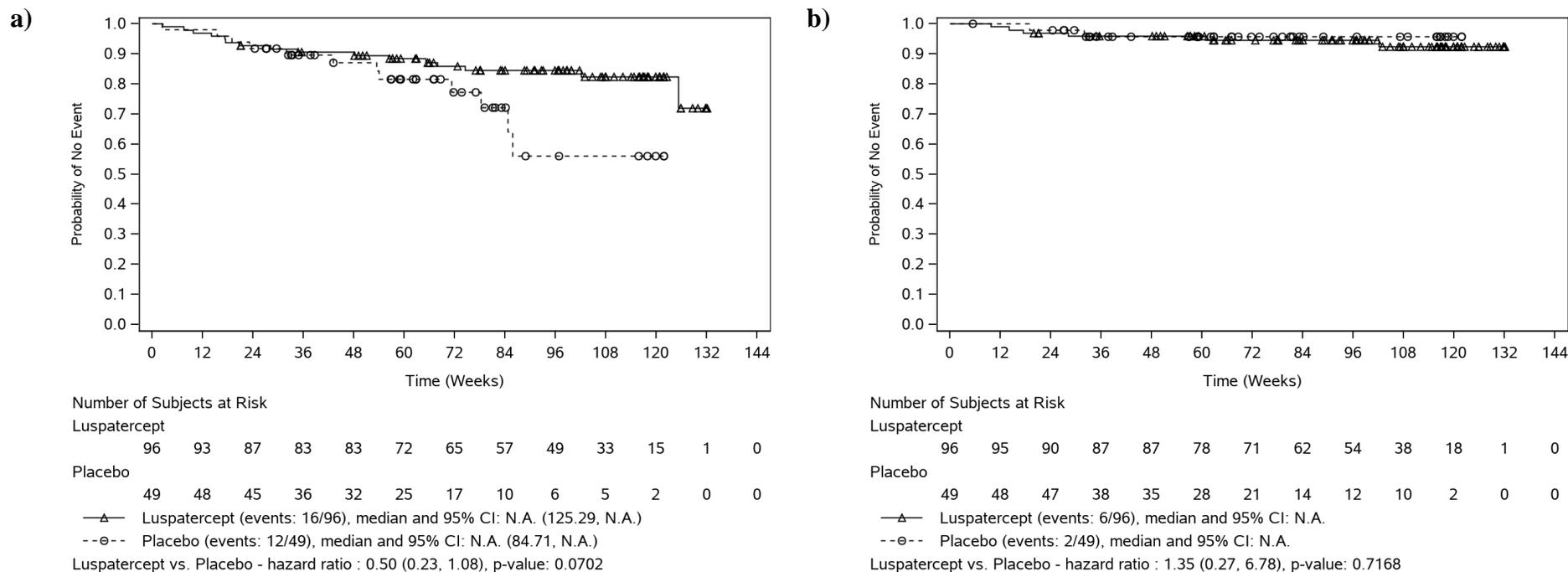


Abbildung 4-18: Kaplan-Meier Plot der Zeit bis zur ersten Hospitalisierung a) unabhängig von der Ursache und b) ohne Berücksichtigung von Hospitalisierungen aufgrund von UE

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

Tabelle 4-60: Ergebnisse für den Endpunkt „Sicherheit“– Gründe für eine Hospitalisierung aus der Studie BEYOND

Studie BEYOND Gründe für eine Hospitalisierung	Luspatercept + BSC <sup>(1)</sup> (N=96)	Placebo + BSC <sup>(1)</sup> (N=49)
	n (%)	n (%)
<b>Gründe für eine Hospitalisierung in Woche 1 – 48</b>		
UE	11 (11,5)	11 (22,4)
Elektive Behandlung einer vorbestehenden Erkrankung	3 (3,1)	1 (2,0)
Transfusion	0	1 (2,0)
Cholezystektomie	1 (1,0)	0
Chirurgische Entfernung von Pilonidalzysten	1 (1,0)	0
Splenektomie	1 (1,0)	0
Akute Pyelonephritis	0	1 (2,0)
ITT-Population; primärer Datenschnitt vom 14.09.2020 <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. BSC: <i>Best-Supportive-Care</i> ; ITT: Intention-to-Treat; UE: unerwünschtes Ereignis		

In der Analyse zu Hospitalisierungen zeigten sich in der Studie BEYOND keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen (siehe Tabelle 4-59). Hauptgrund für Hospitalisierungen waren UE. Entsprechende UE sind definitionsgemäß schwerwiegend und somit bereits in der Analyse der SUE abgebildet. Andere Ursachen für Hospitalisierungen, wie elektive Behandlungen vorbestehender Erkrankungen, traten lediglich vereinzelt auf (siehe Tabelle 4-60). Weiterhin ist die Relevanz verschiedener Gründe der Hospitalisierungen für die Nutzenbewertung fraglich. Vor diesem Hintergrund werden die Ergebnisse zu Hospitalisierungen im Weiteren nicht zur Bewertung des Zusatznutzens herangezogen.

### **Zusammenfassung der Ergebnisse zur Sicherheit**

Bei den Auswertungen zum Endpunkt Sicherheit zeigten sich für Luspatercept im Vergleich zu Placebo sowohl Vor- als auch Nachteile. So ergab sich unter Luspatercept insgesamt ein signifikant geringeres Risiko für schwerwiegende UE. Diesem steht ein signifikant höheres Risiko für spezifische UE, insbesondere bei den PT „Knochenschmerzen“ und „Hypertonie“ gegenüber. Knochenschmerzen und Hypertonie sind für Luspatercept bekannte unerwünschte Arzneimittelwirkungen und entsprechend in der Fachinformation von Reblozyl® aufgeführt [3]. Das häufigere Auftreten von Hypertonie unter Luspatercept wird in den ersten 8 bis 12 Monaten beobachtet und kann eine Anpassung oder Verzögerung der Luspatercept-Dosis oder eine Behandlung gemäß den Empfehlungen der aktuellen klinischen Leitlinien für Hypertonie erfordern. Das häufigere Auftreten von Knochenschmerzen unter Luspatercept ist hierbei auf die ersten 3 Monate der Behandlung begrenzt. Die meisten entsprechenden UE unter Luspatercept (32 von 35 Ereignissen) waren von leichter bis moderater Schwere (Grad 1 oder 2) und nur 3 UE wurden als „schwer“ eingestuft (Grad 3). In der klinischen Praxis sind die anfänglichen Knochenschmerzen, sofern erforderlich, mit Analgetika gut therapierbar. Kein Patient und keine Patientin brach die Behandlung mit Luspatercept aufgrund von Knochenschmerzen ab. Das zeitlich begrenzte Auftreten von Knochenschmerzen deutet auf eine verstärkte Knochenmarksaktivität (Stimulation der Erythropoese), d. h. auf die gewünschte Wirkung von Luspatercept, hin.

Insgesamt zeigt sich somit im Hinblick auf unerwünschte Ereignisse ein mindestens ausgewogenes Bild für Luspatercept im Vergleich zu Placebo, das sich weiterhin mit dem bekannten, gut tolerierbaren Sicherheits- und Verträglichkeitsprofil von Luspatercept deckt.

*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.*

### **Zusammenfassung der Ergebnisse mithilfe von Meta-Analysen**

Nicht zutreffend. Es liegen Ergebnisse einer RCT vor, sodass keine Meta-Analysen durchgeführt wurden.

### **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist in Abschnitt 4.3.1.2.1 beschrieben.

Für den Endpunkt „Sicherheit“ liegt keine abweichende Bewertung vor. Insgesamt ist somit von einer guten Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext auszugehen.

#### 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.<sup>18</sup>

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.

---

<sup>18</sup> unbesetzt

*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.*

Gemäß statistischem Analyseplan der Studie BEYOND waren Subgruppenanalysen hinsichtlich ausgewählter Merkmale u. a. für den primären und die sekundären Endpunkte festgelegt. Für die vorliegende Nutzenbewertung wurden Subgruppenanalysen unter Berücksichtigung dieser präspezifizierten Merkmale für die Hauptanalysen sämtlicher patientenrelevanter Endpunkte vorgenommen.

Auf eine Analyse nach Länder- oder Zentrumseffekten wurde verzichtet, da aussagekräftige Ergebnisse aufgrund von geringen Patientenzahlen pro Studienzentrum bzw. pro Land nicht zu erwarten sind. Stattdessen wurden Subgruppenanalysen zu den übergeordneten, geografischen Regionen (Nordamerika/Europa, Asien-Pazifik und Mittlerer Osten) durchgeführt.

Eine Übersicht der durchgeführten Subgruppenanalysen auf Endpunktebene findet sich in Tabelle 4-61.

Tabelle 4-61: Matrix der durchgeführten Subgruppenanalysen

<b>Studie BEYOND Endpunkt</b>	<b>Alter</b>	<b>Geschlecht</b>	<b>Region</b>	<b>Beta-Thalassämie- Diagnose</b>	<b>Vorhergehende Splenektomie</b>	<b>Hb-Wert zu Baseline</b>	<b>LIC zu Baseline</b>	<b>NTDT-PRO-Gesamtscore T/W zu Baseline</b>	<b>6MWT zu Baseline</b>	<b>FACIT-Fatigue Skala zu Baseline</b>
<b>Mortalität</b>										
Todesfälle	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
<b>Morbidität</b>										
Vermeidung von Transfusionen	○	○	○	○	○	○	○	○	○	○
Symptomatik gemäß NTDT- PRO	●	●	●	●	●	●	●	●	●	●
Symptomatik gemäß PGI-S und PGI-C	○	○	○	○	○	○	○	○	○	○
Körperliche Leistungsfähigkeit gemäß 6MWT	●	●	●	●	●	●	●	●	●	●

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

Studie BEYOND Endpunkt	Alter	Geschlecht	Region	Beta-Thalassämie- Diagnose	Vorhergehende Splenektomie	Hb-Wert zu Baseline	LIC zu Baseline	NTDT-PRO-Gesamtscore T/W zu Baseline	6MWT zu Baseline	FACIT-Fatigue Skala zu Baseline
<b>Lebensqualität</b>										
SF-36	○	○	○	○	○	○	○	○	○	○
FACIT-F	●	●	●	●	●	●	●	●	●	●
<b>Sicherheit</b>										
UE <sup>(1)</sup>	●	●	○	○	●	○	○	○	○	○
<p>●: A priori geplante Subgruppenanalyse. ○: Posthoc durchgeführte Subgruppenanalyse. n. d.: Subgruppenanalyse nicht durchgeführt (Endpunkt mit insgesamt weniger als 10 Ereignissen).</p> <p><sup>(1)</sup> Umfasst Auswertungen zu jeglichen UE, SUE, schweren UE und zum Therapieabbruch führenden UE, zu signifikanten UE auf Ebene der SOC und PT, sowie zu den UE von besonderem Interesse und weiteren untersuchen UE.</p> <p>6MWT: 6-Minuten-Gehtest; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy–Fatigue</i>; Hb: Hämoglobin; LIC: <i>liver iron concentration</i>; NTDT-PRO (T/W bzw. Sob): <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> (Gesamtscore in der Domäne Müdigkeit/Schwäche); PGI-C/S: <i>Patient Global Impression of Change/Severity</i>; SF-36 = <i>Short Form (36) Health Survey</i>; SUE = schwerwiegende UE; UE = unerwünschte Ereignisse</p>										

Stellen Sie anschließend in Tabelle 4-62 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.

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

Tabelle 4-62: Ergebnisse des Interaktionstests der Subgruppenanalysen je Endpunkt

Studie BEYOND Endpunkt	Interaktionstest <sup>(1)</sup>									
	Alter (≤ 32 Jahre vs. > 32 Jahre)	Geschlecht (Weiblich vs. Männlich)	Region (Nordamerika/ Europa vs. Asien-Pazifik vs. Mittlerer Osten)	Beta-Thalassämie- Diagnose (Beta- vs. HbE/Beta-Thalassämie <sup>(2)</sup> )	Vorhergehende Splentektomie (Ja vs. Nein)	Hb-Wert zu Baseline (≥ 8,5 g/dl vs. < 8,5 g/dl)	LIC zu Baseline (< 3 vs. 3 – 5 vs. > 5 – 7 vs. > 7 – 15 vs. > 15 mg/g d. w.)	NTDT-PRO T/W-Gesamt- score zu Baseline (≥ 3 vs. < 3 Punkte)	6MWT zu Baseline (≤ 450 vs. > 450 m)	FACIT-Fatigue Skala zu Baseline (≥ 43 vs. < 43 Punkte)
<b>Morbidität – Vermeidung von Transfusionen</b>										
Vermeidung von Transfusionen	0,5650	0,5638	0,5833	0,7229	0,9856	0,9456	0,5613	0,2816	0,7264	0,0783
<b>Morbidität – Symptomatik gemäß NTDT-PRO</b>										
NTDT-PRO T/W-Gesamtscore - Verbesserung	0,2704	0,0662	0,4390	0,2857	0,5557	<b>0,0354*</b>	0,7242	0,5706	0,6651	0,7916
NTDT-PRO T/W-Gesamtscore - Verschlechterung	0,4683	0,2264	0,0624	0,7283	0,1171	0,3930	0,7355	0,5559	<b>0,0376*</b>	0,3269
NTDT-PRO SoB-Gesamtscore - Verbesserung	0,4904	0,8116	0,9148	0,7506	0,1178	0,0751	0,9907	0,5651	0,3678	0,4695
NTDT-PRO SoB-Gesamtscore - Verschlechterung	0,3626	0,5300	0,5030	0,7632	<b>0,0321*</b>	0,5049	0,9581	0,6250	0,6187	0,5515
<b>Morbidität – Symptomatik gemäß PGI-S und PGI-C</b>										
PGI-S - Verbesserung	0,1453	0,6485	0,3313	0,4443	0,0853	0,1261	0,8467	0,9889	0,7712	0,6698
PGI-S - Verschlechterung	0,6912	0,3790	0,2902	0,9986	0,3545	0,2146	0,5037	0,8578	0,8543	<b>0,0117*</b>
PGI-C - Verbesserung	0,1475	0,0914	0,0615	0,3116	0,3551	0,5073	0,6153	0,8634	0,1863	0,8223
PGI-C - Verschlechterung	0,9940	0,5590	0,6961	0,2897	0,4054	0,3456	0,9965	0,9931	0,1574	0,9999

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

Studie BEYOND Endpunkt	Interaktionstest <sup>(1)</sup>									
	Alter (≤ 32 Jahre vs. > 32 Jahre)	Geschlecht (Weiblich vs. Männlich)	Region (Nordamerika/ Europa vs. Asien-Pazifik vs. Mittlerer Osten)	Beta-Thalassämie- Diagnose (Beta- vs. HbE/Beta-Thalassämie <sup>(2)</sup> )	Vorhergehende Splenektomie (Ja vs. Nein)	Hb-Wert zu Baseline (≥ 8,5 g/dl vs. < 8,5 g/dl)	LIC zu Baseline (< 3 vs. 3 – 5 vs. > 5 – 7 vs. > 7 – 15 vs. > 15 mg/g d. w.)	NTDT-PRO T/W-Gesamt- score zu Baseline (≥ 3 vs. < 3 Punkte)	6MWT zu Baseline (≤ 450 vs. > 450 m)	FACIT-Fatigue Skala zu Baseline (≥ 43 vs. < 43 Punkte)
<b>Morbidität – Körperliche Leistungsfähigkeit gemäß 6MWT</b>										
Körperliche Leistungsfähigkeit gemäß 6MWT	0,5738	0,9841	<b>0,0434*</b>	0,1086	0,6844	0,6800	0,3498	<b>0,0491*</b>	0,6943	0,1076
<b>Lebensqualität – SF-36</b>										
PCS - Verbesserung	0,3098	0,6919	0,0501	0,2225	0,6468	0,5247	0,1430	0,9909	0,0971	0,9913
PCS - Verschlechterung	0,4020	0,8347	0,7337	0,4872	0,2546	0,8738	0,3479	0,7951	0,4454	0,5035
MCS - Verbesserung	0,5096	0,2897	0,9951	0,7074	0,8543	0,8412	0,7764	0,5924	0,1199	0,3774
MCS - Verschlechterung	0,3557	0,5059	0,5382	0,9996	0,7811	0,5366	0,6882	0,5300	0,3287	0,6003
<b>Lebensqualität – FACIT-F</b>										
FACIT-F Gesamtscore - Verbesserung	0,2084	0,9092	0,3385	0,2659	0,2619	0,2289	0,1520	0,3065	0,9892	0,9918
FACIT-F Gesamtscore - Verschlechterung	0,5379	0,7718	0,3572	0,2631	0,7241	0,8807	0,9003	0,8413	0,2808	0,2705
<b>Sicherheit</b>										
Jegliche UE	0,2252	0,2753	0,4465	0,2173	0,3800	0,3186	0,9212	0,5627	0,2068	0,6729
SUE	0,8462	0,9974	0,3649	0,5626	<b>0,0365*</b>	0,5846	n. d.	0,9517	0,8098	0,8227
Schwere UE	0,5789	0,6844	0,3069	0,1540	<b>0,0293*</b>	0,6812	0,2180	0,8433	0,9260	0,8891

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

Studie BEYOND Endpunkt	Interaktionstest <sup>(1)</sup>									
	Alter (≤ 32 Jahre vs. > 32 Jahre)	Geschlecht (Weiblich vs. Männlich)	Region (Nordamerika/ Europa vs. Asien-Pazifik vs. Mittlerer Osten)	Beta- Thalassämie- Diagnose (Beta- vs. HbE/Beta-Thalassämie <sup>(2)</sup> )	Vorhergehende Splenektomie (Ja vs. Nein)	Hb-Wert zu Baseline (≥ 8,5 g/dl vs. < 8,5 g/dl)	LIC zu Baseline (< 3 vs. 3 – 5 vs. > 5 – 7 vs. > 7 – 15 vs. > 15 mg/g d. w.)	NTDT-PRO T/W-Gesamt- score zu Baseline (≥ 3 vs. < 3 Punkte)	6MWT zu Baseline (≤ 450 vs. > 450 m)	FACIT-Fatigue Skala zu Baseline (≥ 43 vs. < 43 Punkte)
Zum Therapieabbruch führende UE	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Skelettmuskulatur-, Bindegewebs- und Knochen- erkrankungen (SOC)	<b>0,0173*</b>	0,6273	<b>0,0437*</b>	0,5033	0,1796	0,4851	0,4374	0,7564	0,1673	0,9557
Knochenschmerzen (PT)	0,2478	0,8026	0,9762	0,7931	0,7725	0,5481	0,9763	0,9941	0,4817	0,4079
Hypertonie (PT)	0,9942	0,9905	> 0,9999	0,9937	0,9897	0,9903	n. d.	0,9929	0,9921	0,9923
Husten (PT)	0,9930	0,9919	> 0,9999	0,9909	0,9900	n. d.	n. d.	0,9930	0,9929	0,9886
Erkrankungen der Haut und des Unterhautgewebes (SOC)	0,9901	0,1673	0,9919	0,5406	0,1523	0,5649	0,9595	0,9888	0,9386	0,8803
Tonsillitis (PT)	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Schwerwiegende Infektionen und parasitäre Erkrankungen (SOC)	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Schwere Infektionen und parasitäre Erkrankungen (SOC)	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Malignitäten <sup>(3)</sup>	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Prä maligne Erkrankungen <sup>(3)</sup>	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.

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

Studie BEYOND Endpunkt	Interaktionstest <sup>(1)</sup>									
	Alter (≤ 32 Jahre vs. > 32 Jahre)	Geschlecht (Weiblich vs. Männlich)	Region (Nordamerika/ Europa vs. Asien-Pazifik vs. Mittlerer Osten)	Beta-Thalassämie- Diagnose (Beta- vs. HbE/Beta-Thalassämie <sup>(2)</sup> )	Vorhergehende Splenoektomie (Ja vs. Nein)	Hb-Wert zu Baseline (≥ 8,5 g/dl vs. < 8,5 g/dl)	LIC zu Baseline (< 3 vs. 3 – 5 vs. > 5 – 7 vs. > 7 – 15 vs. > 15 mg/g d. w.)	NTDT-PRO T/W-Gesamt- score zu Baseline (≥ 3 vs. < 3 Punkte)	6MWT zu Baseline (≤ 450 vs. > 450 m)	FACIT-Fatigue Skala zu Baseline (≥ 43 vs. < 43 Punkte)
Thromboembolische Ereignisse und Thrombophlebitis <sup>(3)</sup>	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Knochenschmerzen und knochenschmerzähnliche Ereignisse <sup>(3)</sup>	0,7283	0,1617	0,7360	0,5483	0,1133	0,1396	0,8324	0,9923	0,4437	0,3932
Hypertonie <sup>(3)</sup>	0,9942	0,9905	> 0,9999	0,9937	0,9897	0,9903	n. d.	0,9929	0,9921	0,9923
Extramedulläre Hämatopoese <sup>(3)</sup>	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Traumatische Frakturen <sup>(3)</sup>	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
Beinulcera <sup>(3)</sup>	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
n. d.: Subgruppenanalyse nicht durchgeführt, da in keiner Subgruppe mindestens 10 Ereignisse auftraten <sup>(1)</sup> Angegebene Werte sind p-Werte; p-Werte < 0,05 wurden mit einem Stern und durch Fettdruck markiert. <sup>(2)</sup> Die a priori definierte Subgruppe „Beta-Thalassämie kombiniert mit Alpha-Thalassämie“ wurde aufgrund einer zu geringen Patientenzahl (weniger als 10 Patient:innen innerhalb der Subgruppe) bei der Subgruppenanalyse nach Beta-Thalassämie-Diagnose nicht berücksichtigt. <sup>(3)</sup> UE-Kategorien, die im Rahmen der präspezifizierten UE von besonderem Interesse und weiteren untersuchten UE betrachtet wurden (vgl. Abschnitt 4.2.5.4). 6MWT: 6-Minuten-Gehtest; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy-Fatigue</i> ; Hb: Hämoglobin; LIC: <i>liver iron concentration</i> ; MCS: <i>Mental Component Summary</i> ; NTDT-PRO (T/W bzw. SoB): <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> (Gesamtscore in der Domäne Müdigkeit/Schwäche bzw. Kurzatmigkeit); PCS: <i>Physical Component Summary</i> ; PGI-C/S: <i>Patient Global Impression of Change/Severity</i> ; PT: <i>Preferred Term</i> ; SF-36: <i>Short Form (36) Health Survey</i> ; SOC: <i>System Organ Class</i> ; SUE: schwerwiegendes UE; UE: unerwünschtes Ereignis										

*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.*

### **Darstellung der Subgruppenanalysen**

Die Interaktions-p-Werte der Subgruppenanalysen für alle patientenrelevanten Endpunkte finden sich in Tabelle 4-62. Die Ergebnisse der Subgruppenanalysen mit einem Interaktions-p-Wert  $< 0,05$  sind in den folgenden Abschnitten dargestellt.

Subgruppen mit einem Interaktions-p-Wert  $\geq 0,05$  werden als homogen hinsichtlich des betrachteten Behandlungseffekts beurteilt. Die Ergebnisse sind ergänzend in Anhang 4-G dargestellt. Bei 255 im Rahmen der Subgruppenanalysen berechneten Interaktionstests traten lediglich 10 signifikante Interaktions-p-Werte, verteilt auf die Subgruppenfaktoren Alter, Region, vorhergehende Splenektomie, Hb-Wert zu Baseline, NTDT-PRO T/W-Gesamtscore zu Baseline, 6MWT zu Baseline und FACIT-Fatigue Skala zu Baseline, auf. Es handelt sich daher um keine sich wiederholenden (endpunktübergreifende) Effekte. Demensprechend liegen keine belastbaren Hinweise vor, dass sich der Zusatznutzen von Luspatercept zwischen verschiedenen Subgruppen unterscheiden könnte. Die Bewertung des Zusatznutzens (siehe Abschnitt 4.4.2) erfolgt daher ausschließlich basierend auf den Ergebnissen der Gesamtpopulation.

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

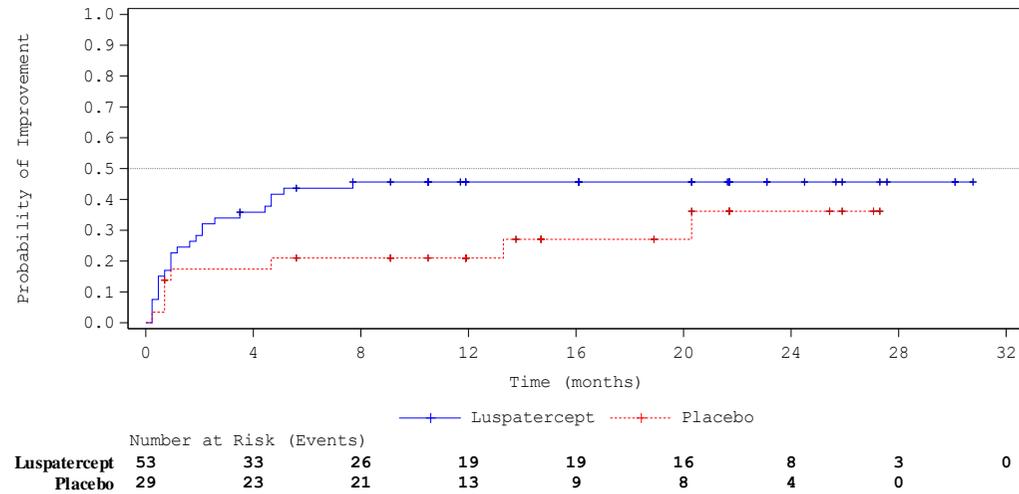
**Subgruppenanalysen für den Endpunkt „Morbidity - Symptomatik gemäß NTDT-PRO“**

Tabelle 4-63: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ T/W-Gesamtscore – Zeit bis zur ersten bewertungsrelevanten Verbesserung aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zur ersten bewertungsrelevanten Verbesserung des NTDT-PRO Gesamtscores in der Domäne T/W								
Hb-Wert zu Baseline – Interaktions-p-Wert 0,0354								
< 8,5 g/dl	53	24 (45,3)	N.A. [3,5; N.A.]	29	8 (27,6)	N.A. [20,3; N.A.]	1,80 [0,80; 4,04]	0,1324
≥ 8,5 g/dl	41	21 (51,2)	9,1 [0,9; N.A.]	20	14 (70,0)	1,2 [0,5; 11,9]	0,62 [0,32; 1,23]	0,1679
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.								
<sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verbesserung								
<sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell mit Adjustierung für den Ausgangswert								
<sup>(4)</sup> Unstratifizierter Log-Rank-Test								
BSC: <i>Best-Supportive-Care</i> ; Hb: Hämoglobin; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; NTDT-PRO T/W: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> Gesamtscore in der Domäne Müdigkeit/Schwäche								

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

a)



b)

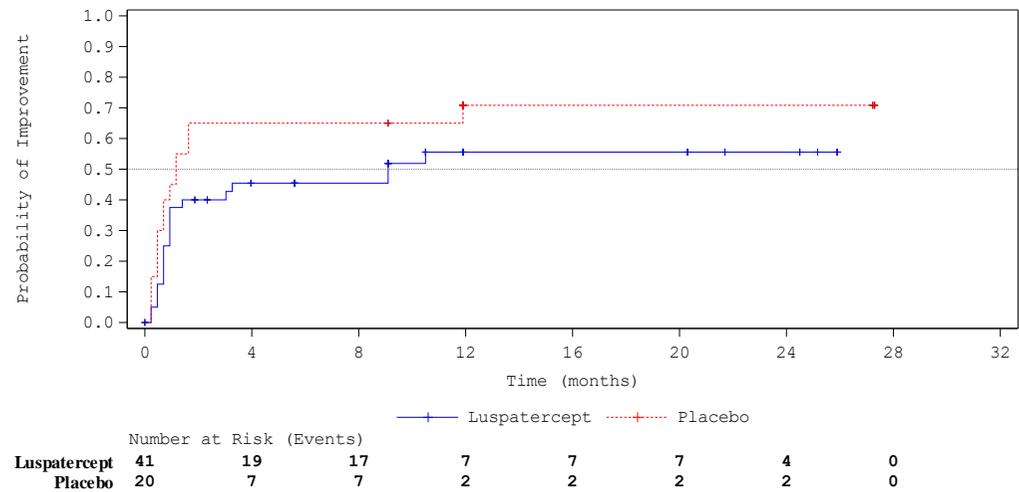


Abbildung 4-19: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verbesserung des NTDT-PRO Gesamtscores in der Domäne T/W für die Subgruppe a) Hb-Wert zu Baseline < 8,5 g/dl und b) Hb-Wert zu Baseline ≥ 8,5 g/dl

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

Tabelle 4-64: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ T/W-Gesamtscore – Zeit bis zur ersten bewertungsrelevanten Verschlechterung aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne T/W								
6MWT zu Baseline – Interaktions-p-Wert 0,0376								
≤ 450 m	50	14 (28,0)	N.A. [27,6; N.A.]	31	6 (19,4)	N.A. [20,3; N.A.]	1,76 [0,67; 4,66]	0,4631
> 450 m	44	5 (11,4)	N.A. [N.A.; N.A.]	18	4 (22,2)	N.A. [2,3; N.A.]	0,39 [0,10; 1,50]	0,2656
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell mit Adjustierung für den Ausgangswert <sup>(4)</sup> Unstratifizierter Log-Rank-Test								
6MWT: 6-Minuten-Gehtest; BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; NTDT-PRO T/W: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> Gesamtscore in der Domäne Müdigkeit/Schwäche								

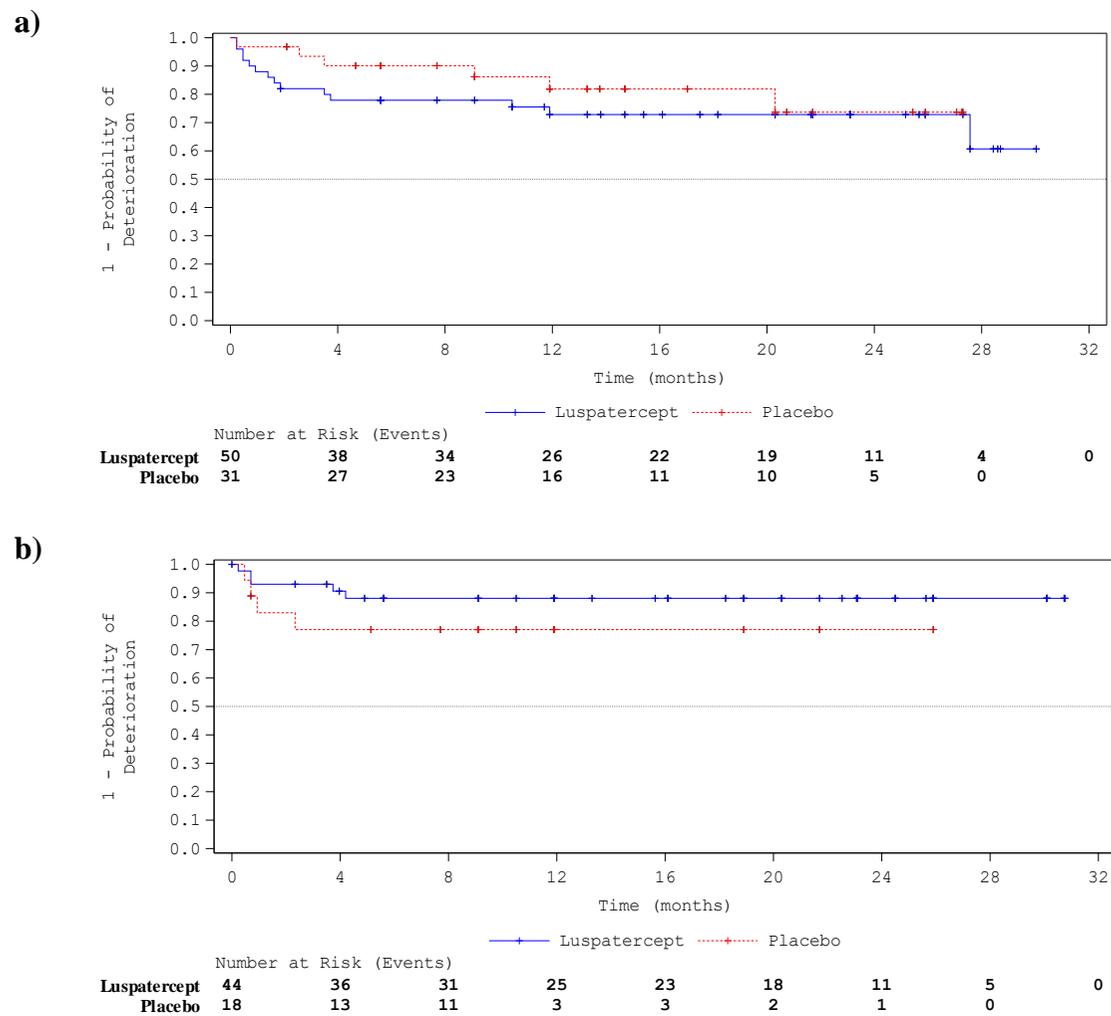


Abbildung 4-20: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne T/W für die Subgruppe a) 6MWT zu Baseline ≤ 450 m und b) 6MWT zu Baseline > 450 m

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

Tabelle 4-65: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß NTDT-PRO“ SoB-Gesamtscore – Zeit bis zur ersten bewertungsrelevanten Verschlechterung aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignissen (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignissen (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne SoB								
Vorhergehende Splenektomie – Interaktions-p-Wert 0,0321								
ja	33	10 (30,3)	N.A. [N.A.; N.A.]	26	3 (11,5)	N.A. [N.A.; N.A.]	2,92 [0,80; 10,65]	0,1049
nein	61	13 (21,3)	N.A. [N.A.; N.A.]	23	9 (39,1)	N.A. [3,5; N.A.]	0,52 [0,22; 1,23]	0,1072
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie. <sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell mit Adjustierung für den Ausgangswert <sup>(4)</sup> Unstratifizierter Log-Rank-Test BSC: <i>Best-Supportive-Care</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; NTDT-PRO SoB: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> Gesamtscore in der Domäne Kurzatmigkeit								

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

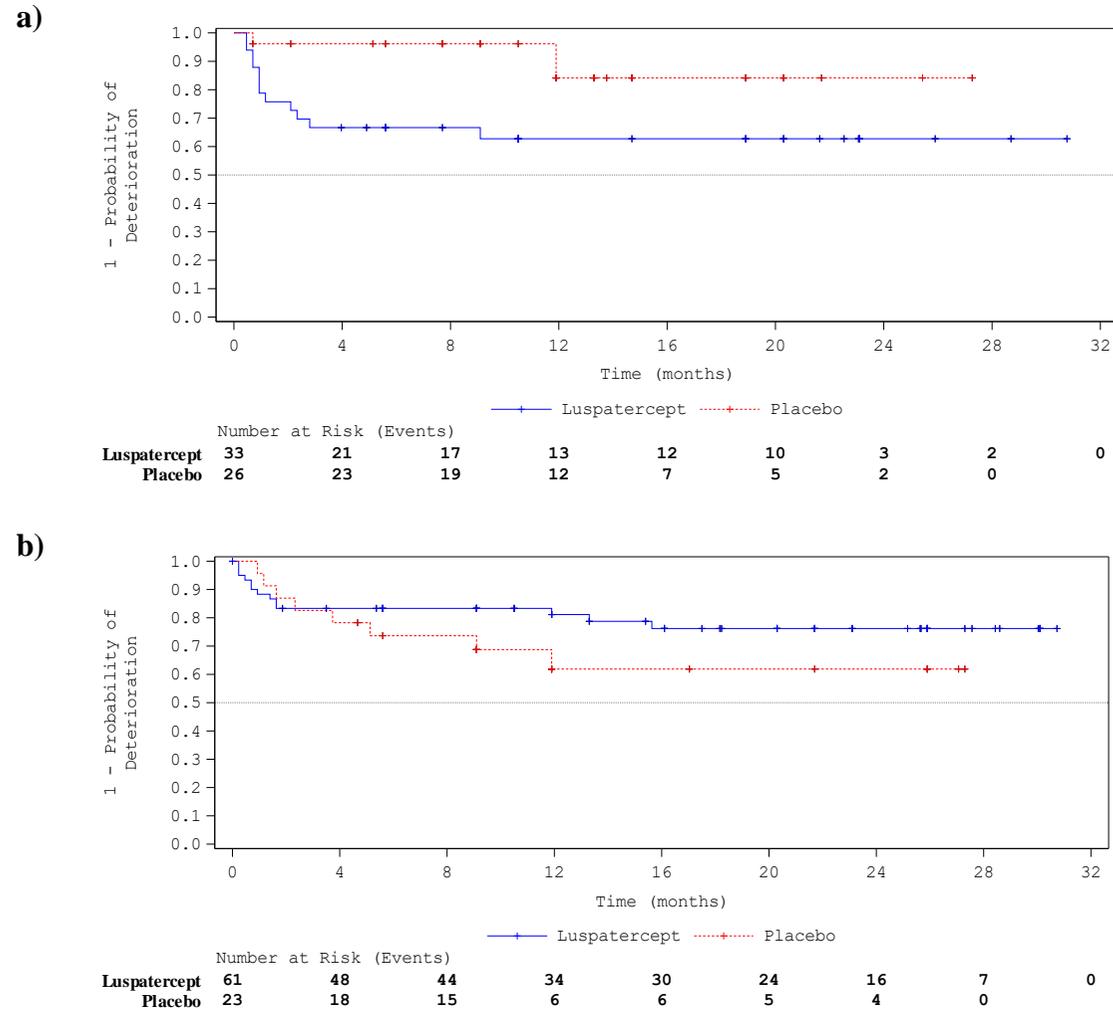


Abbildung 4-21: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne SoB für die Subgruppe a) mit vorhergehender Splenektomie zu Baseline und b) ohne vorhergehende Splenektomie zu Baseline

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

Für den Endpunkt „Morbidity - Symptomatik gemäß NTDT-PRO“ wurden bei der Auswertung der Zeit bis zur ersten bewertungsrelevanten Verbesserung des NTDT-PRO Gesamtscores in der Domäne T/W eine potenzielle Effektmodifikation durch den Hb-Wert zu Baseline (< 8,5 g/dl vs. ≥ 8,5 g/dl) festgestellt, sowie darüber hinaus bei der Auswertung der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne T/W eine potenzielle Effektmodifikation durch den 6MWT zu Baseline (≤ 450 m vs. > 450 m) und bei der Auswertung der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des NTDT-PRO Gesamtscores in der Domäne SoB eine potenzielle Effektmodifikation durch den Subgruppenfaktor vorhergehende Splenektomie (ja vs. nein). Die Effekte innerhalb der Subgruppen sind jedoch – konsistent mit dem Ergebnis in der Gesamtpopulation – durchgängig nicht statistisch signifikant (siehe Tabelle 4-63, Tabelle 4-64 und Tabelle 4-65).

**Subgruppenanalysen für den Endpunkt „Morbidity - Symptomatik gemäß PGI-S“**

Tabelle 4-66: Subgruppenanalysen für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S“ – Zeit bis zur ersten bewertungsrelevanten Verschlechterung aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zur ersten bewertungsrelevanten Verschlechterung des PGI-S								
FACIT-Fatigue Skala zu Baseline – Interaktions-p-Wert 0,0117								
< 43 Punkte	63	19 (30,2)	N.A. [N.A.; N.A.]	32	7 (21,9)	N.A. [20,3; N.A.]	1,60 [0,67; 3,80]	0,3498
≥ 43 Punkte	31	6 (19,4)	N.A. [27,6; N.A.]	16	9 (56,3)	10,2 [0,5; N.A.]	0,21 [0,07; 0,65]	<b>0,0014</b>
ITT-Population; primärer Datenschnitt vom 14.09.2020								
<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.								
<sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zur ersten bewertungsrelevanten Verschlechterung								
<sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell mit Adjustierung für den Ausgangswert								
<sup>(4)</sup> Unstratifizierter Log-Rank-Test								
BSC: <i>Best-Supportive-Care</i> ; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy–Fatigue</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; PGI-S: <i>Patient Global Impression of Severity</i>								

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

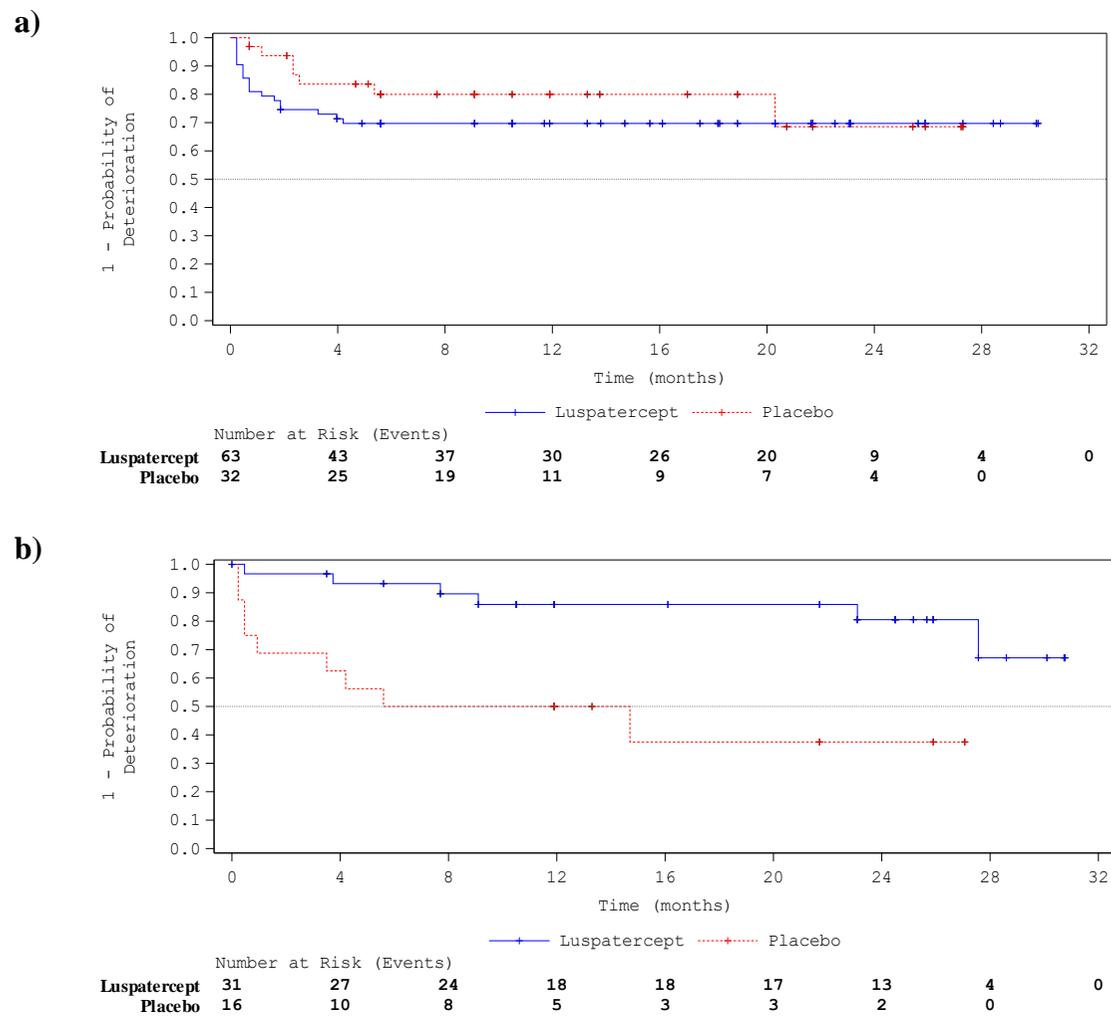


Abbildung 4-22: Kaplan-Meier Plot der Zeit bis zur ersten bewertungsrelevanten Verschlechterung des PGI-S für die Subgruppe a) FACIT-Fatigue Skala zu Baseline < 43 Punkte und b) FACIT-Fatigue Skala zu Baseline ≥ 43 Punkte

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

Für den Endpunkt „Morbidity – Symptomatik gemäß PGI-S“ wurde eine potenzielle Effektmodifikation bei der Auswertung der Zeit bis zur ersten bewertungsrelevanten Verschlechterung durch den Baseline-Score der Fatigue-spezifische Skala des FACIT-F (< 43 Punkte vs. ≥ 43 Punkte) festgestellt. Während sich in der Subgruppe mit höherer Fatigue zu Baseline (FACIT-Fatigue Skala < 43 Punkte) konsistent zum Ergebnis der Gesamtpopulation kein Unterschied zwischen den Studienarmen zeigte, ergab sich in der Subgruppe mit geringerer Fatigue zu Baseline (FACIT-Fatigue Skala ≥ 43 Punkte) ein signifikanter Effekt zugunsten von Luspatercept (siehe Tabelle 4-66).

**Subgruppenanalysen für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)“**

Tabelle 4-67: Subgruppenanalysen für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)“ – Änderung des 6MWT gegenüber Baseline aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel [95 %-KI] <sup>(4)</sup>	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel [95 %-KI] <sup>(4)</sup>	Mittlere Differenz [95 %-KI]; p-Wert <sup>(4)</sup>	Hedges' g [95 %-KI] <sup>(5)</sup>
Mittlere Änderung des 6MWT gegenüber Baseline zu Woche 48								
Region – Interaktions-p-Wert 0,0434								
Nordamerika/Europa	52	461,70 (83,11)	18,92 [6,26, 31,58]	22	449,36 (85,62)	-15,85 [-35,33, 3,63]	34,77 [11,52, 58,02] <b>0,0039</b>	0,65 [0,20, 1,10]
Mittlerer Osten	8	359,50 (170,47)	62,53 [37,28, 87,79]	8	392,00 (81,36)	73,34 [48,08, 98,60]	-10,81 [-46,68, 25,06] 0,5264	-0,27 [-1,23, 0,68]
Asien-Pazifik	26	424,37 (60,40)	-27,93 [-41,78, -14,07]	11	412,00 (39,64)	-17,05 [-38,40, 4,29]	-10,88 [-36,38, 14,63] 0,3923	-0,29 [-1,00, 0,41]
NTDT-PRO T/W-Gesamtscore zu Baseline – Interaktions-p-Wert 0,0491								
< 3 Punkte	27	466,22 (67,94)	-16,73 [-33,12, -0,34]	14	415,07 (74,92)	0,95 [-22,24, 24,14]	-17,68 [-46,82, 11,45] 0,2267	-0,38 [-1,02, 0,26]

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

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>	
	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel [95 %-KI] <sup>(4)</sup>	N <sup>(2)</sup>	Baseline Mittel (SD) <sup>(3)</sup>	Änderung gegenüber Baseline Mittel [95 %-KI] <sup>(4)</sup>	Mittlere Differenz [95 %-KI]; p-Wert <sup>(4)</sup>	Hedges' g [95 %-KI] <sup>(5)</sup>
≥ 3 Punkte	59	429,32 (100,37)	22,45 [8,34, 36,55]	27	434,93 (79,08)	-2,86 [-23,72, 17,99]	25,31 [0,12, 50,50] <b>0,0489</b>	0,41 [0,00, 0,82]

ITT-Population; Primärer Datenschnitt vom 14.09.2020

<sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie.  
<sup>(2)</sup> Patient:innen mit einem 6MWT-Wert zu Baseline und zu Woche 48  
<sup>(3)</sup> Arithmetischer Mittelwert basierend auf dem letzten Wert vor Therapiebeginn  
<sup>(4)</sup> ANCOVA basierend auf der Änderung gegenüber Baseline, adjustiert für den Baseline-6MWT-Wert  
<sup>(5)</sup> Hedges' g abgeleitet aus dem Modellschätzer der mittleren Differenz

ANCOVA: *analysis of covariance*; BSC: *Best-Supportive-Care*; Hb: Hämoglobin; ITT: Intention-to-Treat; KI: Konfidenzintervall; SD: Standardabweichung; SE: Standardfehler

Für den Endpunkt „Morbidity – Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)“ wurde bei der Auswertung zur Änderung des 6MWT gegenüber Baseline eine potenzielle Effektmodifikation durch die geographische Region (Nordamerika/Europa vs. Asien-Pazifik vs. Mittlerer Osten) und durch den NTDT-PRO T/W-Gesamtscore zu Baseline (< 3 Punkte vs. ≥ 3 Punkte) festgestellt.

Hinsichtlich der potenziellen Effektmodifikation durch die geographische Region zeigte sich in der Region Nordamerika/Europa ein signifikanter Effekt zugunsten von Luspatercept, während sich in den Regionen Mittlerer Osten und Asien-Pazifik konsistent zur Gesamtpopulation kein Unterschied zwischen den Behandlungsarmen ergab (siehe Tabelle 4-67).

Hinsichtlich der potenziellen Effektmodifikation durch den NTDT-PRO T/W-Gesamtscore zu Baseline zeigte sich in der Subgruppe mit stärkerer Müdigkeit und Schwäche zu Baseline (NTDT-PRO T/W-Gesamtscore ≥ 3 Punkte) ein signifikanter Effekt zugunsten von Luspatercept, während sich in der Subgruppe mit geringerer Müdigkeit und Schwäche zu Baseline (NTDT-PRO T/W-Gesamtscore < 3 Punkte) konsistent zur Gesamtpopulation kein Unterschied zwischen den Behandlungsarmen ergab (siehe Tabelle 4-67).

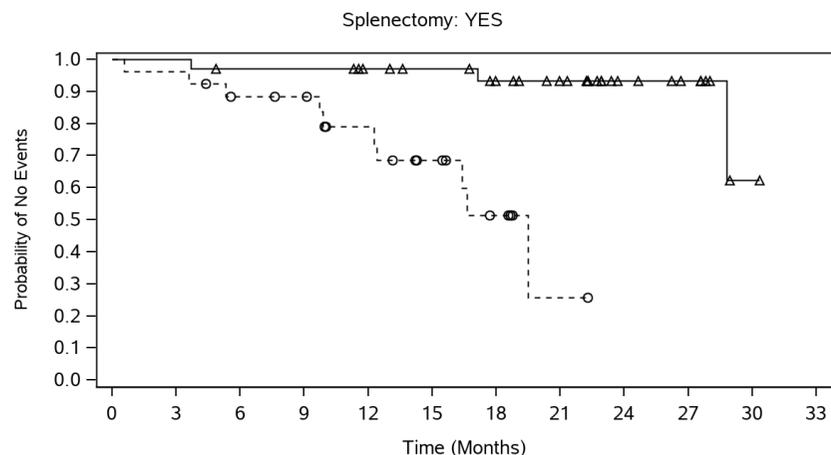
Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

**Subgruppenanalysen für den Endpunkt „Sicherheit“**

Tabelle 4-68: Subgruppenanalysen für den Endpunkt „Sicherheit“ – Zeit bis zum ersten SUE aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zum ersten SUE								
Vorhergehende Splenektomie – Interaktions-p-Wert 0,0365								
Ja	34	3 (8,8)	N.A. [28,81; N.A.]	26	10 (38,5)	19,48 [12,42; N.A.]	0,08 [0,02; 0,37]	<b>&lt; 0,0001</b>
Nein	62	8 (12,9)	N.A. [N.A.; N.A.]	23	2 (8,7)	N.A. [18,00; N.A.]	1,18 [0,25; 5,62]	0,8315
Safety-Population; primärer Datenschnitt vom 14.09.2020 <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie. <sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zum ersten SUE <sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell <sup>(4)</sup> Unstratifizierter Log-Rank-Test BSC: <i>Best-Supportive-Care</i> ; CTCAE: <i>Common Terminology Criteria for Adverse Events</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; SUE: schwerwiegendes unerwünschtes Ereignis								

a)

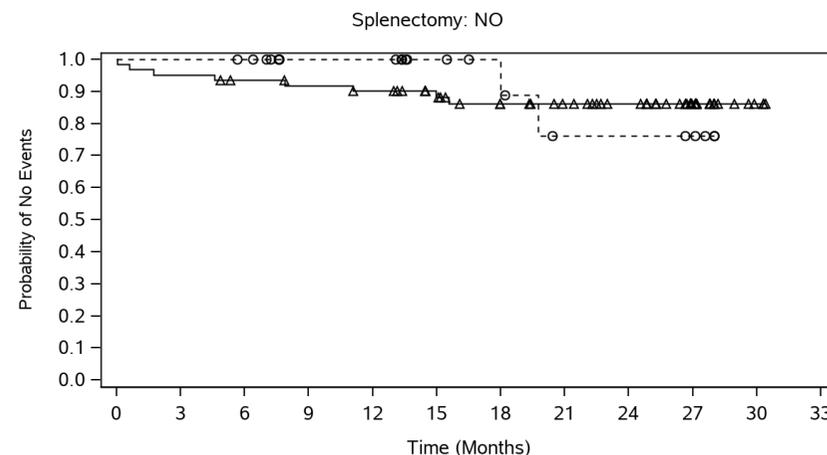


Number of Subjects at Risk

Luspatercept	34	34	32	32	29	27	23	19	10	7	1	0
Placebo	26	25	21	20	15	10	5	1	0	0	0	0

—▲— Luspatercept (events: 3/34), median and 95% CI: N.A. (28.81, N.A.)  
 - -○- - Placebo (events: 10/26), median and 95% CI: 19.48 (12.42, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 0.08 (0.02, 0.37), pvalue: <0.0001

b)



Number of Subjects at Risk

Luspatercept	62	59	56	54	52	46	39	35	29	16	3	0
Placebo	23	23	22	16	16	11	9	5	5	4	0	0

—▲— Luspatercept (events: 8/62), median and 95% CI: N.A.  
 - -○- - Placebo (events: 2/23), median and 95% CI: N.A. (18.00, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 1.18 (0.25, 5.62), pvalue: 0.8315

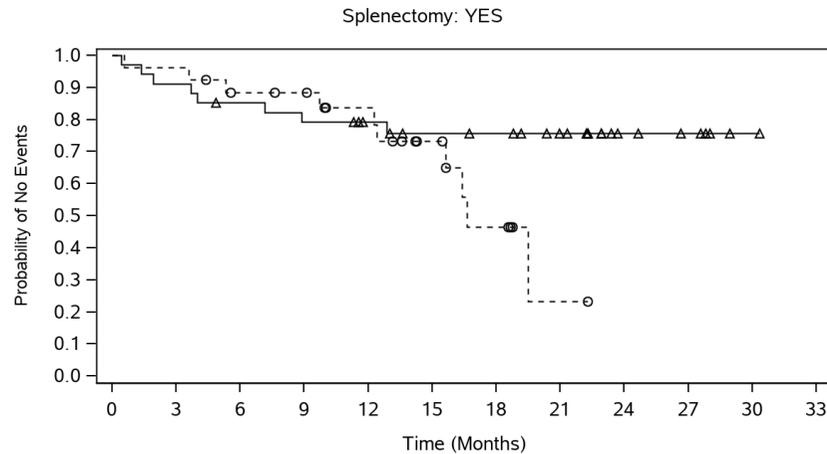
Abbildung 4-23: Kaplan-Meier Plot der Zeit bis zum ersten SUE für die Subgruppe a) mit vorhergehender Splenektomie und b) ohne vorhergehende Splenektomie

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

Tabelle 4-69: Subgruppenanalysen für den Endpunkt „Sicherheit“ – Zeit bis zum ersten schweren UE aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zum ersten schweren UE (CTCAE-Grad ≥ 3)								
Vorhergehende Splenektomie – Interaktions-p-Wert 0,0293								
Ja	34	8 (23,5)	N.A. [N.A.; N.A.]	26	10 (38,5)	16,62 [12,42; N.A.]	0,48 [0,18; 1,25]	0,1241
Nein	62	19 (30,6)	N.A. [N.A.; N.A.]	23	2 (8,7)	N.A. [18,00; N.A.]	3,38 [0,79; 14,54]	0,0824
Safety-Population; primärer Datenschnitt vom 14.09.2020 <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelattherapie. <sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zum ersten UE <sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell <sup>(4)</sup> Unstratifizierter Log-Rank-Test BSC: <i>Best-Supportive-Care</i> ; CTCAE: <i>Common Terminology Criteria for Adverse Events</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar								

a)

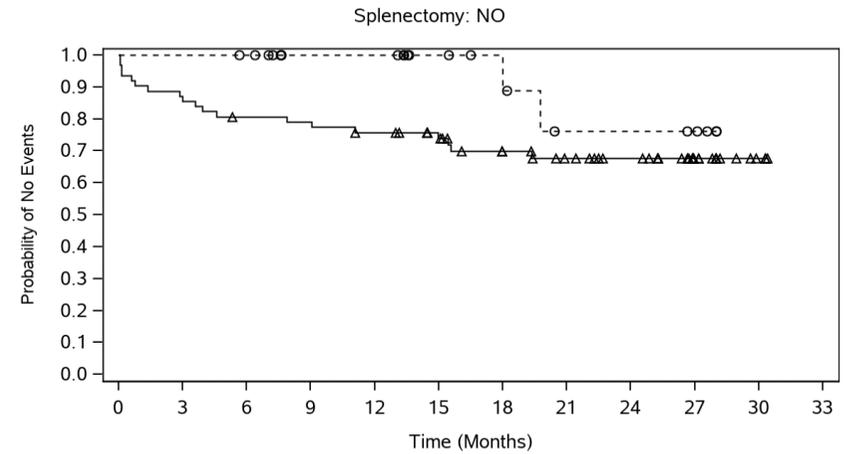


Number of Subjects at Risk

Luspatercept	34	31	28	26	23	20	19	15	7	5	1	0
Placebo	26	25	21	20	16	10	5	1	0	0	0	0

—▲— Luspatercept (events: 8/34), median and 95% CI: N.A.  
 - -○- - Placebo (events: 10/26), median and 95% CI: 16.62 (12.42, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 0.48 (0.18, 1.25), pvalue: 0.1241

b)



Number of Subjects at Risk

Luspatercept	62	53	49	48	45	40	32	27	22	12	3	0
Placebo	23	23	22	16	16	11	9	5	5	4	0	0

—▲— Luspatercept (events: 19/62), median and 95% CI: N.A.  
 - -○- - Placebo (events: 2/23), median and 95% CI: N.A. (18.00, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 3.38 (0.79, 14.54), pvalue: 0.0824

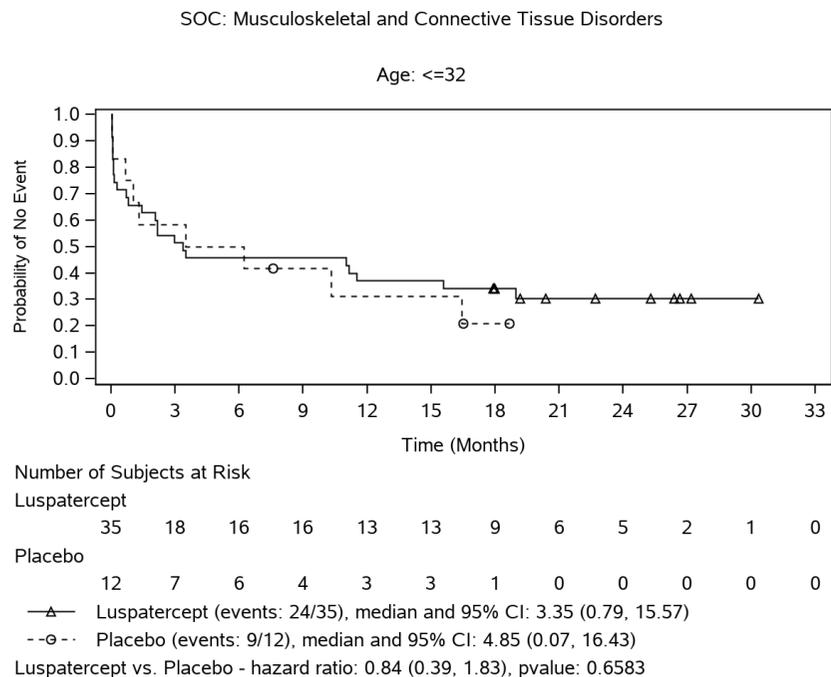
Abbildung 4-24: Kaplan-Meier Plot der Zeit bis zum ersten schweren UE für die Subgruppe a) mit vorhergehender Splenektomie und b) ohne vorhergehende Splenektomie

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

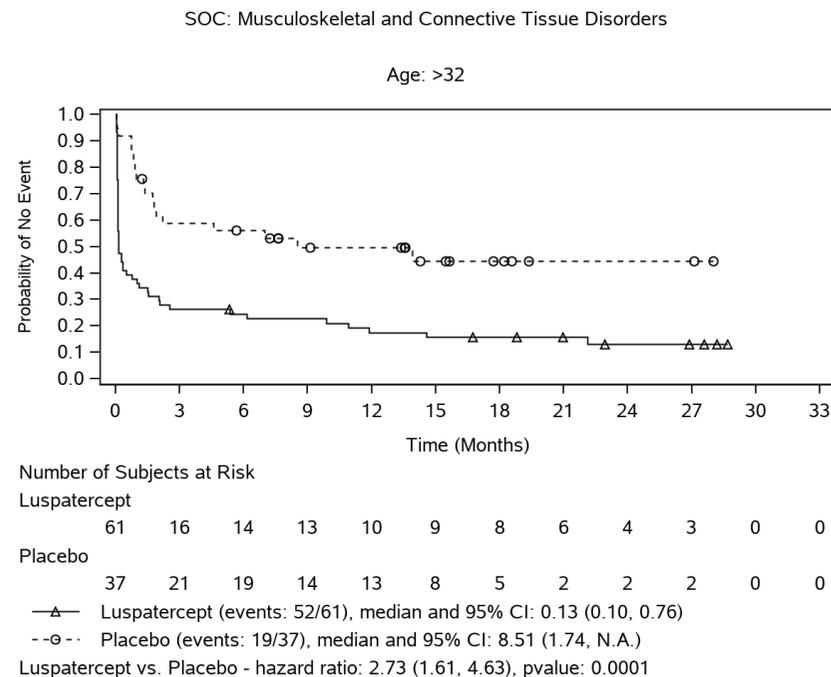
Tabelle 4-70: Subgruppenanalysen für den Endpunkt „Sicherheit“ – Zeit bis zum ersten UE nach SOC Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen aus der Studie BEYOND

Studie BEYOND	Luspatercept + BSC <sup>(1)</sup> (N=96)			Placebo + BSC <sup>(1)</sup> (N=49)			Luspatercept + BSC vs. Placebo + BSC <sup>(1)</sup>	
	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	N	Patient:innen mit Ereignis n (%)	Median (Monate) [95 %-KI] <sup>(2)</sup>	Hazard Ratio [95 %-KI] <sup>(3)</sup>	p-Wert <sup>(4)</sup>
Zeit bis zum ersten UE der SOC Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen								
Alter – Interaktions-p-Wert 0,0173								
≤ 32 Jahre	35	24 (68,6)	3,35 [0,79; 15,57]	12	9 (75,0)	4,85 [0,07; 16,43]	0,84 [0,39; 1,83]	0,6583
> 32 Jahre	61	52 (85,2)	0,13 [0,10; 0,76]	37	19 (51,4)	8,51 [1,74; N.A.]	2,73 [1,61; 4,63]	<b>0,0001</b>
Region – Interaktions-p-Wert 0,0437								
Nordamerika/Europa	60	48 (80,0)	0,30 [0,10; 2,07]	30	14 (46,7)	N.A. [1,38; N.A.]	2,49 [1,37; 4,53]	<b>0,0025</b>
Mittlerer Osten	9	7 (77,8)	2,17 [0,03; N.A.]	8	8 (100,0)	0,77 [0,03; 8,51]	0,57 [0,21; 1,59]	0,2816
Asien-Pazifik	27	21 (77,8)	0,79 [0,10; 10,91]	11	6 (54,5)	13,90 [0,13; N.A.]	2,09 [0,84; 5,21]	0,1067
Safety-Population; primärer Datenschnitt vom 14.09.2020 <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Kaplan-Meier Schätzer der medianen Zeit bis zum ersten UE <sup>(3)</sup> Unstratifiziertes Cox-Regressionsmodell <sup>(4)</sup> Unstratifizierter Log-Rank-Test BSC: <i>Best-Supportive-Care</i> ; CTCAE: <i>Common Terminology Criteria for Adverse Events</i> ; KI: Konfidenzintervall; N.A.: nicht anwendbar, nicht erreicht oder nicht berechenbar; SOC: <i>System Organ Class</i> ; UE = unerwünschtes Ereignis								

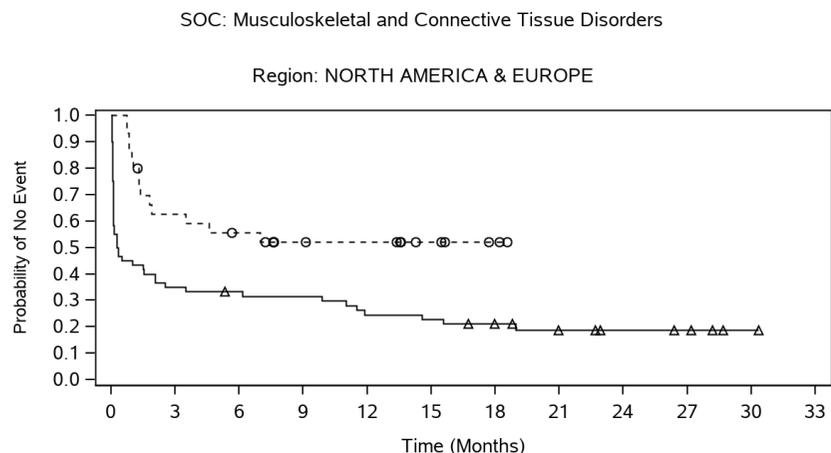
a)



b)



c)

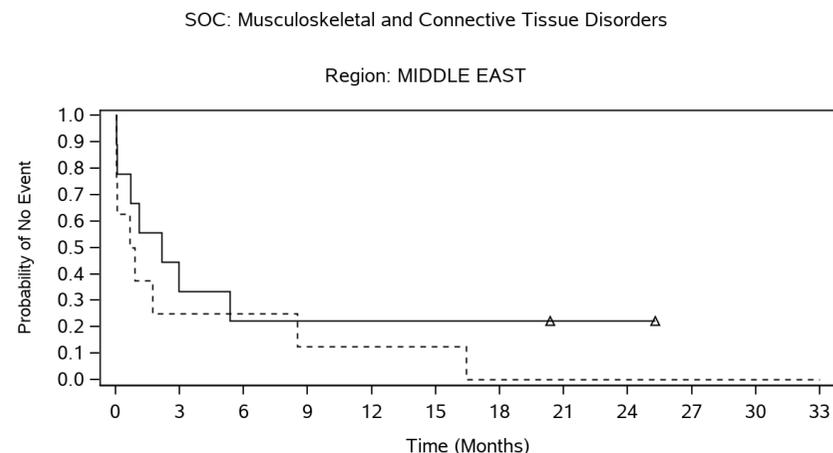


Number of Subjects at Risk

Luspatercept	60	21	19	18	14	13	10	7	5	4	1	0
Placebo	30	18	15	10	9	5	2	0	0	0	0	0

—▲— Luspatercept (events: 48/60), median and 95% CI: 0.30 (0.10, 2.07)  
 - -○- - Placebo (events: 14/30), median and 95% CI: N.A. (1.38, N.A.)  
 Luspatercept vs. Placebo - hazard ratio: 2.49 (1.37, 4.53), pvalue: 0.0025

d)



Number of Subjects at Risk

Luspatercept	9	3	2	2	2	2	2	1	1	0	0	0
Placebo	8	2	2	1	1	1	0	0	0	0	0	0

—▲— Luspatercept (events: 7/9), median and 95% CI: 2.17 (0.03, N.A.)  
 - -○- - Placebo (events: 8/8), median and 95% CI: 0.77 (0.03, 8.51)  
 Luspatercept vs. Placebo - hazard ratio: 0.57 (0.21, 1.59), pvalue: 0.2816

e)

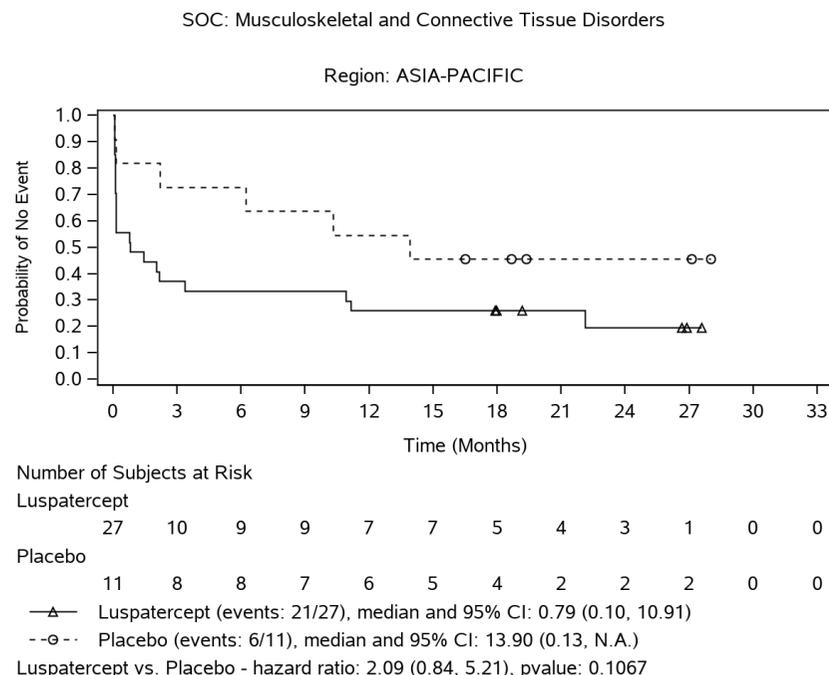


Abbildung 4-25: Kaplan-Meier Plot der Zeit bis zum ersten UE der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen“ für die Subgruppe a) Alter ≤ 32 Jahre, b) Alter > 32 Jahre, c) Region Nordamerika/Europa, d) Region Mittlerer Osten und e) Region Asien-Pazifik

Für den Endpunkt „Sicherheit“ wurde bei der Auswertung der Zeit bis zum ersten SUE und bei der Auswertung der Zeit bis zum ersten schweren UE jeweils eine potenzielle Effektmodifikation durch den Subgruppenfaktor vorhergehende Splenektomie (ja vs. nein) festgestellt. Darüber hinaus wurde bei der Auswertung der Zeit bis zum ersten UE der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen“ jeweils eine potenzielle Effektmodifikation durch das Alter der Patient:innen (≤ 32 Jahre vs. > 32 Jahre) und durch die geographische Region (Nordamerika/Europa vs. Asien-Pazifik vs. Mittlerer Osten) festgestellt.

---

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

Hinsichtlich der potenziellen Effektmodifikation durch eine vorhergehende Splenektomie bei den SUE zeigte sich in der Subgruppe mit vorhergehender Splenektomie konsistent zum Ergebnis der Gesamtpopulation ein signifikanter Effekt zugunsten von Luspatercept, während sich in der Subgruppe ohne vorhergehende Splenektomie kein Unterschied zwischen den Behandlungsarmen ergab (siehe Tabelle 4-68).

Hinsichtlich der potenziellen Effektmodifikation durch eine vorhergehende Splenektomie bei den schweren UE zeigte sich in beiden Subgruppen konsistent zum Ergebnis der Gesamtpopulation kein Unterschied zwischen den Behandlungsarmen (siehe Tabelle 4-69).

Hinsichtlich der potenziellen Effektmodifikation durch das Alter bei den UE der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen“ zeigte sich in der Subgruppe der älteren Patient:innen konsistent zum Ergebnis der Gesamtpopulation ein signifikanter Effekt zuungunsten von Luspatercept, während sich in der Subgruppe der jüngeren Patient:innen kein Unterschied zwischen den Behandlungsarmen ergab (siehe Tabelle 4-70).

Hinsichtlich der potenziellen Effektmodifikation durch die geographische Region bei den UE der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen“ zeigte sich in der Region Nordamerika/Europa konsistent zum Ergebnis der Gesamtpopulation ein signifikanter Effekt zuungunsten von Luspatercept, während sich in den Regionen Mittlerer Osten und Asien-Pazifik kein Unterschied zwischen den Behandlungsarmen ergab (siehe Tabelle 4-70).

#### 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 BEYOND

- Publikationen [41]
- Studienberichte [42]
- Studienregistereinträge [39, 40]

#### 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 4-71: 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 4-72: 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 4-73: 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 4-74: Bewertung des Verzerrungspotenzials für &lt;Endpunkt xxx&gt; 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 4-75: Ergebnisse für &lt;Endpunkt xxx&gt; 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 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 4-76: 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 4-77: 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 4-78: 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 4-79: Verzerrungsaspekte für &lt;Endpunkt xxx&gt; – 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 4-80: 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 4-81: 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.*

Zur Darstellung des Zusatznutzens von Luspatercept bei Erwachsenen für die Behandlung von Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist, wurde im vorliegenden Dossier die Zulassungsstudie BEYOND herangezogen.

Die Studie BEYOND ist eine doppelblinde, randomisiert kontrolliert durchgeführte, multi-zentrische Studie, in der die Sicherheit und Wirksamkeit von Luspatercept gegenüber Placebo, jeweils in Kombination mit BSC, untersucht wurde. In beiden Behandlungsarmen war im Rahmen von BSC eine bedarfsgerechte Transfusionstherapie mit EK in Kombination mit Eisenchelatoren erlaubt, sofern angezeigt. Eine bedarfsgerechte Transfusionstherapie mit EK in Kombination mit Eisenchelatoren entspricht der vom G-BA im Beratungsgespräch festgelegten zVT [4].

Die Studie BEYOND wurde doppelblind durchgeführt. Die Aufrechterhaltung der Verblindung erfolgte unter anderem durch eine zur Luspatercept-Therapie analoge Placebo-Gabe. Gemäß § 5 Absatz 6 des 5. Kapitels der Verfo entspricht die Studie BEYOND als RCT der Evidenzstufe 1b. Somit ist die Studie unter der Voraussetzung statistisch signifikanter Ergebnisse bei Endpunkten mit einem niedrigen Verzerrungspotenzial geeignet, um eine Zusatznutzenaussage auf Ebene eines Hinweises abzuleiten.

Im vorliegenden Dossier wurde zum Zwecke einer konservativen Einschätzung für unerwünschte Ereignisse sowie für die patientenberichteten Endpunkte zur Lebensqualität und zur Symptomatik gemäß NTDT-PRO ein erhöhtes Verzerrungspotenzial abgeleitet. Demgegenüber liegt jedoch für die Mortalität sowie die Morbiditätsendpunkte „Vermeidung von Transfusionen“, „Symptomatik gemäß PGI-S und PGI-C“ und „Körperliche Leistungsfähigkeit gemäß 6MWT“ jeweils ein niedriges Verzerrungspotenzial vor. Folglich können Aussagen zum Zusatznutzen auf Basis der Ergebnisse dieser Endpunkte als „Hinweis“ eingestuft werden.

#### **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.*

Die vorliegende Nutzenbewertung erfolgt auf Basis der randomisierten, kontrollierten Doppelblindstudie BEYOND. Die demographischen und krankheitsspezifischen Charakteristika der eingeschlossenen Patient:innen sowie das Studiendesign legen eine gute Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext nahe.

Tabelle 4-82 fasst die Ergebnisse der patientenrelevanten Endpunkte aus der Studie BEYOND zusammen (Hauptanalysen).

Tabelle 4-82: Zusammenfassung der Therapieeffekte von Luspatercept

Studie BEYOND Endpunkt	Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		Ausmaß des Zusatz- nutzens	
	Effektschätzer [95 %-KI]	p-Wert		
<b>Mortalität</b>				
Todesfälle	N.A. (Anteil: 0 % vs. 0 %)	↔	N.A.	Nicht belegt
<b>Morbidität</b>				
<b>Vermeidung von Transfusionen</b> Zeit bis zur ersten Transfusion	HR = 0,23 [0,11; 0,46]	↑	< 0,0001	<b>Beträchtlicher Zusatznutzen</b>
<b>Regulation der Erythropoese (ergänzend dargestellt)</b> Änderung des Hb (g/dl) bis Woche 48	MWD = 1,35 [1,12; 1,57]	↑	< 0,0001	
<b>Kontrolle der Eisenüberladung (ergänzend dargestellt)</b> Änderung des sFerritin (µg/l) bis Woche 48	MWD = 13,50 [-60,96; 87,96]	↔	0,7205	
<b>Symptomatik gemäß NTDT-PRO</b> <i>Gesamtscore in der Domäne Müdigkeit/Schwäche</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 1,03 [0,61; 1,71]	↔	0,8972	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,94 [0,43; 2,04]	↔	0,8670	
<i>Gesamtscore in der Domäne Kurzatmigkeit</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 1,20 [0,67; 2,14]	↔	0,2348	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 1,00 [0,49; 2,01]	↔	0,9714	
<b>Symptomatik gemäß PGI-S und PGI-C</b> <i>PGI-S</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 2,28 [1,15; 4,50]	↑	<b>0,0244</b>	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,71 [0,38; 1,32]	↔	0,2191	
<i>PGI-C</i> Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 6,44 [3,19; 12,97]	↑	< 0,0001	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,51 [0,15; 1,71]	↔	0,2686	
<b>Körperliche Leistungsfähigkeit</b> Änderung im 6-Minuten Gehstest (Meter) zu Woche 48	MWD = 12,44 [-6,72; 31,59]	↔	0,2011	

Studie BEYOND Endpunkt	Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		Ausmaß des Zusatz- nutzens
	Effektschätzer [95 %-KI]	p-Wert	
<b>Gesundheitsbezogene Lebensqualität</b>			
<b>SF-36 – PCS</b>			<b>Beträchtlicher Zusatznutzen</b>
Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 2,79 [1,06; 7,38]	↑ <b>0,0457</b>	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,66 [0,29; 1,50]	↔ 0,3364	
<b>SF-36 – MCS</b>			
Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 1,44 [0,68; 3,08]	↔ 0,1056	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,59 [0,33; 1,07]	↑ <b>0,0303</b>	
<b>FACIT-F – Gesamtscore</b>			
Zeit bis zur ersten bewertungsrelevanten Verbesserung <sup>(2)</sup>	HR = 2,41 [0,98; 5,95]	↑ <b>0,0340</b>	
Zeit bis zur ersten bewertungsrelevanten Verschlechterung <sup>(2)</sup>	HR = 0,78 [0,38; 1,59]	↔ 0,5099	
<b>Sicherheit</b>			
<b>UE-Hauptkategorien – Zeit bis zum ersten Ereignis</b>			<b>Nicht belegt</b>
<u>Schwerwiegende</u> UE (SUE)	HR = 0,29 [0,12; 0,69]	↑ <b>0,0032</b>	
<u>Schwere</u> UE (CTCAE ≥ 3)	HR = 1,07 [0,54; 2,14]	↔ 0,8419	
Zum Therapieabbruch führende UE	HR = 0,29 [0,06; 1,34]	↔ 0,0923	
<b>Signifikante Unterschiede bei <u>jeglichen</u> UE auf Ebene der SOC und PT – Zeit bis zum ersten Ereignis</b>			
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen (SOC)	HR = 1,85 [1,19; 2,86]	↓ <b>0,0065</b>	
Knochenschmerzen (PT)	HR = 7,11 [2,18; 23,15]	↓ <b>0,0001</b>	
Tonsillitis (PT)	HR = 0,13 [0,03; 0,64]	↑ <b>0,0033</b>	
Hypertonie (PT)	HR = 9,32 [1,25; 69,66]	↓ <b>0,0079</b>	
Husten (PT)	HR = 7,24 [0,96; 54,68]	↓ <b>0,0248</b>	
Erkrankungen der Haut und des Unterhautgewebes (SOC)	HR = 2,59 [0,99; 6,77]	↓ <b>0,0434</b>	
<b>Signifikante Unterschiede bei <u>schwer- wiegenden/schweren</u> UE auf Ebene der SOC und PT – Zeit bis zum ersten Ereignis</b>			
<u>Schwerwiegende</u> Infektionen und parasitäre Erkrankungen (SOC)	HR = 0,11 [0,02; 0,57]	↑ <b>0,0017</b>	
<u>Schwere</u> Infektionen und parasitäre Erkrankungen (SOC)	HR = 0,05 [0,01; 0,42]	↑ <b>0,0002</b>	

Studie BEYOND Endpunkt	Luspatercept + BSC <sup>(1)</sup> vs. Placebo + BSC <sup>(1)</sup>		Ausmaß des Zusatz- nutzens
	Effektschätzer [95 %-KI]	p-Wert	
Primärer Datenschnitt: 14.09.2020 ↑ Signifikanter Effekt zugunsten von Luspatercept; ↔ Kein signifikanter Effekt; ↓ Signifikanter Effekt zuungunsten von Luspatercept <sup>(1)</sup> BSC umfasste insbesondere die bedarfsgerechte Transfusionstherapie mit Erythrozytenkonzentraten in Kombination mit einer Eisenchelatherapie. <sup>(2)</sup> Eine bewertungsrelevante Verbesserung/Verschlechterung entspricht einer Responseschwelle von 15 % der Skalen-spannweite bzw. im Falle des PGI-C den beiden besten/schlechtesten Bewertungen der Veränderung. BSC: <i>Best Supportive Care</i> ; FACIT-F: <i>Functional Assessment of Chronic Illness Therapy–Fatigue</i> ; HR: <i>Hazard Ratio</i> ; MCS: <i>Mental Component Summary</i> ; MWD: <i>Mittelwertdifferenz</i> ; NTDT-PRO: <i>Nontransfusion-Dependent Thalassemia-Patient Reported Outcomes</i> ; PCS: <i>Physical Component Summary</i> ; PGI-C/S: <i>Patient Global Impression of Change/Severity</i> ; PT: <i>Preferred Term</i> ; RR: <i>relatives Risiko</i> SF-36: <i>Short Form (36) Health Survey</i> ; SOC: <i>System Organ Class</i> ; SUE: <i>schwerwiegende UE</i> ; UE: <i>unerwünschte Ereignisse</i>			

Bezogen auf die einzelnen Endpunktkategorien ist das Ausmaß des Zusatznutzens für Luspatercept zur Behandlung von erwachsenen Patient:innen mit Anämie, die mit einer nicht-transfusionsabhängigen Beta-Thalassämie verbunden ist, maßgeblich wie folgt begründet:

### Mortalität

In der doppelblinden Studienphase der Studie BEYOND verstarben keine Patient:innen.

In der Endpunktkategorie Mortalität ist ein Zusatznutzen von Luspatercept gegenüber der zVT somit nicht belegt.

### Morbidität

In der Endpunktkategorie Morbidität wurde der Zusatznutzen anhand der patientenrelevanten Endpunkte *Vermeidung von Transfusionen*, *Veränderung in der Symptomatik* (NTDT-PRO, PGI-S und PGI-C) und *körperliche Belastbarkeit* (6-Minuten-Gehtest) bewertet. Klinisch relevante, therapiesteuernde Parameter der Anämie (Hb-Werte) und Eisenüberladung (sFerritin), die im vorliegenden Anwendungsgebiet einen besonderen Stellenwert haben, werden ergänzend berichtet (siehe Tabelle 4-82).

### *Vermeidung von Transfusionen und Linderung der Anämie*

In der Ereigniszeitanalyse der Zeit bis zur ersten Transfusion ergab sich für den Vergleich von Luspatercept gegenüber Placebo ein HR von 0,23 (95 %-KI [0,11; 0,46];  $p < 0,0001$ ). Dies entspricht einer statistisch signifikanten **Reduktion des Risikos für Transfusionen um 77 %**. Auch in den zusätzlichen Responderanalysen zur Transfusionsfreiheit über 24 und 48 Wochen ergab sich jeweils ein statistisch signifikanter Vorteil für den Luspatercept-Arm gegenüber dem Placebo-Arm.

Die Vermeidung von Transfusionen unter Luspatercept ging einher mit einer Steigerung und langfristigen Stabilisierung der Hb-Werte und somit einer anhaltenden Linderung der Anämie. Im Durchschnitt ergab sich über die Behandlungsphase bis Woche 48 im Luspatercept-Arm ein mittlerer Anstieg der Hb-Werte um 1,44 g/dl während es im Placebo-Arm nur 0,10 g/dl waren (mittlere Differenz: 1,35 [1,12; 1,57];  $p < 0,0001$ ). Die Erhöhung des mittleren Hb-Spiegels im Luspatercept-Arm wurde dazu über den gesamten Behandlungsverlauf bis zum betrachteten Datenschnitt stabil aufrechterhalten.

Die Vermeidung von Transfusionen und Linderung der Anämie stellen eine beträchtliche Verbesserung des therapielevanten Nutzens dar, die mit einer klinisch relevanten und für die Patient:innen spürbaren Verbesserung ihres Zustands einhergeht:

- Erstens profitieren die Patient:innen langfristig von eingesparten EK-Transfusionen aufgrund der Vermeidung von Transfusionsrisiken, Folgekomplikationen sowie einer weiteren Eisenüberladung.
- Zweitens ist vor dem chronischen Hintergrund der NTDT die Vermeidung von Transfusionen sowie die Verhinderung eines Übergangs in eine transfusionsabhängige Beta-Thalassämie, die mit regelmäßiger Transfusionstherapie und einem erhöhten Behandlungsaufwand, insbesondere einem zeitlichen Aufwand, einhergeht, von großer Bedeutung für die Betroffenen.
- Drittens führt die durch Luspatercept bedingte Steigerung und Stabilisierung der Hb-Werte (Verbesserung der Erythropoese) nicht nur zu einer langanhaltenden Vermeidung von Transfusionen, sondern ermöglicht im Weiteren auch eine Reduktion Anämie-assoziiertes Symptomatik und folglich auch eine spürbare Verbesserung im Befinden der Patient:innen.

Diese therapiebedingten Verbesserungen der Erkrankung spiegeln sich auch klar in der patientenberichteten Veränderung der Symptomatik und der Lebensqualität wider.

### ***Symptomatik gemäß NTDT-PRO***

Im krankheitsspezifischen Fragebogen NTDT-PRO wurden im Hinblick auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung und Verschlechterung (Hauptanalysen) unter Luspatercept jeweils keine signifikanten Vorteile gegenüber Placebo beobachtet. Demgegenüber zeigten sich in den Zusatzanalysen signifikante Vorteile für Luspatercept: So betrug der Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung des NTDT-PRO-Gesamtscores in der Domäne Müdigkeit/Schwäche zu Woche 24 im Luspatercept-Arm 35,5 % gegenüber 17,9 % im Placebo-Arm ( $p = 0,0434$ ). Dieser signifikante Vorteil setzte sich zu Woche 48 mit 41,7 % gegenüber 11,1 % fort ( $p = 0,0317$ ). Darüber hinaus betrug der Anteil der Patient:innen mit einer bewertungsrelevanten Verbesserung des NTDT-PRO-Gesamtscores in der Domäne Kurzatmigkeit zu Woche 24 im Luspatercept-Arm 27,6 % gegenüber 10,3 % im Placebo-Arm ( $p = 0,0332$ ). Dieser Vorteil blieb zu Woche 48 mit 35,4 % gegenüber 11,1 % numerisch erhalten ( $p = 0,0594$ ).

Entsprechende Unterschiede sind auch in den Abbildungen der mittleren Verläufe und den MMRM-Analysen ersichtlich (siehe Abschnitt 4.3.1.3.1.4 und Anhang 4-G). Zudem lassen die Abbildungen vermuten, dass die Vorteile im Rahmen der Ereigniszeitanalysen aufgrund von vorübergehenden Anfangsverbesserungen im Placebo-Arm ggf. unzureichend abgebildet werden.

Insgesamt sind die Ergebnisse zum NTDT-PRO positiv zu bewerten. Die Zusatzanalysen deuten auf eine spürbare Verbesserung der krankheitsspezifischen Symptomatik unter Luspatercept hin.

### ***Symptomatik gemäß PGI-S und PGI-C***

Für die patientenberichtete Symptomatik gemäß PGI-S und PGI-C zeigte sich in Bezug auf die Zeit bis zur ersten Verbesserung jeweils ein sehr deutlicher, statistisch signifikanter Vorteil für Luspatercept im Vergleich zu Placebo.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des PGI-S ergab sich ein signifikanter Effekt mit einem HR von 2,28 (95 %-KI [1,15; 4,50];  $p = 0,0244$ ), d. h. die **Chance auf eine bewertungsrelevante Verbesserung des PGI-S war unter Luspatercept mehr als doppelt so hoch** wie unter Placebo.

In der Ereigniszeitanalyse zur ersten bewertungsrelevanten Verbesserung des PGI-C ergab sich ebenfalls ein signifikanter Effekt mit einem HR von 6,44 (95 %-KI [3,19; 12,97];  $p < 0,0001$ ), d. h. die **Chance auf eine bewertungsrelevante Verbesserung des PGI-C war unter Luspatercept mehr als 6-mal höher** als unter Placebo.

Diese Ergebnisse wurden durch die Zusatzanalysen zum Anteil der Patient:innen mit einer signifikanten Verbesserung zu Woche 24 und 48 bestätigt und spiegeln sich darüber hinaus auch eindeutig in den Ergebnissen der MMRM-Analysen und den Verlaufskurven zum PGI-S bzw. in der Abbildung zu den PGI-C-Bewertungen im Studienverlauf wider (siehe Abschnitt 4.3.1.3.1.5 und Anhang 4-G).

Die Ergebnisse zum PGI-S und PGI-C verdeutlichen, dass die anhaltende Linderung der Anämie entsprechend der Steigerung und langfristigen Stabilisierung der Hb-Werte durch Luspatercept auch zu einer sehr deutlichen und patientenrelevanten Reduktion der durch die Patient:innen wahrgenommenen Symptombelastung führt.

### ***Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest***

In Bezug auf die körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest zeigte sich in der Studie BEYOND kein Unterschied zwischen den Behandlungsarmen.

### ***Fazit Morbidität***

Zusammengenommen führt die Behandlung mit Luspatercept im Vergleich zur zVT zu bedeutenden Vorteilen in Bezug auf die langfristige Vermeidung von Transfusionen in Verbindung mit einer anhaltenden Linderung der Anämie sowie einer spürbaren und deutlichen Reduktion der Krankheitssymptomatik. **Für die Endpunktkategorie Morbidität besteht daher ein beträchtlicher Zusatznutzen von Luspatercept gegenüber der zVT.**

### **Lebensqualität**

In der Endpunktkategorie gesundheitsbezogene Lebensqualität wurde der Zusatznutzen in der Studie BEYOND anhand eines generischen Fragebogens (SF-36) und eines Fatigue-spezifischen Fragebogens (FACIT-F) bewertet. Die durch die ineffektive Erythropoese bedingte Anämie kann die Lebensqualität von Beta-Thalassämie-Patient:innen maßgeblich beeinträchtigen (siehe auch Modul 3 C). Die Aufrechterhaltung oder Verbesserung der Lebensqualität ist im vorliegenden Anwendungsgebiet daher ein wichtiges Therapieziel.

#### ***SF-36***

Für den Endpunkt SF-36 zeigten sich in der Studie BEYOND im Rahmen der Ereigniszeitanalysen statistisch signifikante Effekte zugunsten von Luspatercept in Bezug auf die Zeit bis zur ersten bewertungsrelevanten Verbesserung des PCS (körperliches Wohlbefinden) und die Zeit bis zur ersten bewertungsrelevanten Verschlechterung des MCS (psychisches Wohlbefinden).

Für den PCS ergab sich bei der Zeit bis zur ersten bewertungsrelevanten Verbesserung ein HR von 2,79 (95 %-KI [1,06; 7,38];  $p = 0,0457$ ), d. h. die **Chance auf eine Verbesserung des körperlichen Wohlbefindens war unter Luspatercept mehr als doppelt so hoch** wie unter Placebo.

Für den MCS ergab sich bei der Zeit bis zur ersten bewertungsrelevanten Verschlechterung ein HR von 0,59 (95 %-KI [0,33; 1,07];  $p = 0,0303$ ), d. h. **das Risiko für eine Verschlechterung des psychischen Wohlbefindens war unter Luspatercept 41 % geringer** als unter Placebo.

#### ***FACIT-F***

Für den Endpunkt FACIT-F zeigte sich in der Studie BEYOND bei der Zeit bis zur Verbesserung des Gesamtscores ein statistisch signifikanter Effekt zugunsten von Luspatercept.

Für den FACIT-F Gesamtscore ergab sich bei der Zeit bis zur ersten bewertungsrelevanten Verbesserung ein HR von 2,41 (95 %-KI [0,98; 5,95];  $p = 0,0340$ ), d. h. die **Chance auf eine Verbesserung des Gesamtscores der Fatigue-spezifischen Lebensqualität war unter Luspatercept mehr als doppelt so hoch** wie unter Placebo.

### ***Fazit Lebensqualität***

Zusammengenommen geht aus den Ergebnissen hervor, dass die anhaltende Linderung der Anämie unter Luspatercept auch zu einer spürbaren Verbesserung der Lebensqualität der NTDT-Patient:innen führt. **Für die Endpunktkategorie gesundheitsbezogene Lebensqualität besteht daher ein beträchtlicher Zusatznutzen von Luspatercept gegenüber der zVT.**

### **Sicherheit**

Die Auswertung der unerwünschten Ereignisse (UE) in der Studie BEYOND zeigt, dass die Behandlung mit Luspatercept im Vergleich zu Placebo insgesamt mit einer signifikanten Reduktion des Risikos für schwerwiegende UE (SUE) sowie jeweils für schwere und schwerwiegende UE der SOC „Infektionen und parasitäre Erkrankungen“ einherging.

Darüber hinaus ergaben sich auf SOC- und PT-Ebene keine weiteren Unterschiede bei schwerwiegenden und schweren UE. Auch bei zum Therapieabbruch führenden UE wurden keine Unterschiede festgestellt.

Bei Betrachtung jeglicher UE auf Ebene der SOC und PT zeigte sich ein Vorteil bei dem PT „Tonsillitis“, während bei der SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“, dem PT „Knochenschmerzen“, dem PT „Hypertonie“, dem PT „Husten“ und der SOC „Erkrankungen der Haut und des Unterhautgewebes“ signifikante Unterschiede zuungunsten von Luspatercept festgestellt wurden.

Knochenschmerzen und Hypertonie sind für Luspatercept bekannte unerwünschte Arzneimittelwirkungen und entsprechend in der Fachinformation von Reblozyl® aufgeführt [3]. Das häufigere Auftreten von Hypertonie unter Luspatercept wird in den ersten 8 bis 12 Monaten beobachtet und kann eine Anpassung oder Verzögerung der Luspatercept-Dosis oder eine Behandlung gemäß den Empfehlungen der aktuellen klinischen Leitlinien für Hypertonie erfordern. Das häufigere Auftreten von Knochenschmerzen unter Luspatercept ist auf die ersten 3 Monate begrenzt. Die meisten entsprechenden UE unter Luspatercept (32 von 35 Ereignissen) waren von leichter bis moderater Schwere (Grad 1 oder 2) und nur 3 UE wurden als „schwer“ eingestuft (Grad 3). In der klinischen Praxis sind die anfänglichen Knochenschmerzen, sofern erforderlich, mit Analgetika gut therapierbar. Kein Patient und keine Patientin brach die Behandlung mit Luspatercept aufgrund von Knochenschmerzen ab. Das zeitlich begrenzte Auftreten von Knochenschmerzen deutet auf eine verstärkte Knochenmarksaktivität (Stimulation der Erythropoese), d. h. auf die gewünschte Wirkung von Luspatercept, hin

### ***Fazit Sicherheit***

Insgesamt besteht im Hinblick auf UE ein mindestens ausgewogenes Bild für Luspatercept im Vergleich zu Placebo, das sich weiterhin mit dem bekannten, gut tolerierbaren Sicherheits- und Verträglichkeitsprofil von Luspatercept deckt. Aufgrund des höheren Risikos für spezifische Ereignisse (insbesondere Knochenschmerzen und Hypertonie) wird trotz der signifikanten Reduktion der Gesamtrate schwerwiegender UE für die Endpunktkategorie Sicherheit kein Zusatznutzen von Luspatercept gegenüber der zVT abgeleitet.

## Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen

### Zusatznutzen

Die Beta-Thalassämie ist eine seltene, erblich bedingte, hämatologische Erkrankung mit einer erheblichen Krankheitslast und Beeinträchtigung der Lebensqualität. Dabei spielen sowohl erkrankungsbedingte als auch therapiebedingte (z. B. transfusionsbedingte klinische und psychosoziale) Komplikationen eine große Rolle. Im Zentrum der Komplikationen steht die Anämie infolge der ineffektiven Erythropoese aufgrund des vorliegenden Erythrozyten-Reifungs-Defekts.

Für Patient:innen im Anwendungsgebiet von Luspatercept gibt es einen großen therapeutischen Bedarf nach einem Arzneimittel, das den Erythrozyten-Reifungs-Defekt adressiert und mit einer Erhöhung und Stabilisierung des Hb-Wertes die Anämie lindert. NTDT-Patient:innen leiden häufig über ausgedehnte Zeiträume unter niedrigen Hb-Werten sowie entsprechenden Anämie-assoziierten Symptomen und erfahren nur selten eine Linderung. Zudem treten Folgekomplikationen, bedingt durch die Grunderkrankung, häufiger und mitunter schwerwiegender bei NTDT-Patient:innen auf, da diese im Gegensatz zu den TDT-Patient:innen keine kontinuierliche Transfusionstherapie zur Unterdrückung der eigenen ineffektiven Erythropoese erhalten. Aber auch die Entwicklung einer regelmäßigen Transfusionsnotwendigkeit und damit ein Wechsel des klinischen Bildes hin zu dem einer TDT ist möglich.

Luspatercept ist der erste Vertreter der neuen Wirkstoffklasse der **Erythrozyten-Reifungs-Aktivatoren** (ERA) und damit die erste aktive, medikamentöse Therapie zur Behandlung der Anämie bei Erwachsenen mit nicht-transfusionsabhängiger Beta-Thalassämie. Die Therapie mit Luspatercept adressiert gezielt den der Anämie zugrundeliegenden Erythrozyten-Reifungs-Defekt. Dies führt zu einer Steigerung und langfristigen Stabilisierung der Hb-Werte, wodurch einerseits Transfusionen und potenzielle Folgekomplikationen vermieden werden können und andererseits Symptomatik, Lebensqualität und Autonomie der Patient:innen verbessert werden. Damit deckt Luspatercept den großen therapeutischen Bedarf im vorliegenden Anwendungsgebiet.

Die Ergebnisse zum medizinischen Nutzen und Zusatznutzen von Luspatercept beruhen auf der doppelblinden, randomisierten, kontrollierten Zulassungsstudie BEYOND. Gemäß § 5 Absatz 6 des 5. Kapitels der VerfO entspricht die Studie BEYOND als RCT der Evidenzstufe 1b. Somit ist die Studie auf Basis signifikanter Ergebnisse bei Endpunkten mit einem niedrigen Verzerrungspotenzial geeignet, um eine Zusatznutzenaussage auf Ebene eines Hinweises abzuleiten. Im vorliegenden Dossier wurde zum Zwecke einer konservativen Einschätzung für unerwünschte Ereignisse sowie für die patientenberichteten Endpunkte zur Lebensqualität und zur Symptomatik gemäß NTDT-PRO ein erhöhtes Verzerrungspotenzial abgeleitet. Demgegenüber liegt allerdings u. a. für die Morbiditätsendpunkte „Vermeidung von Transfusionen“ und „Symptomatik gemäß PGI-S und PGI-C“ jeweils ein niedriges Verzerrungspotenzial vor. Folglich können Aussagen zum Zusatznutzen auf Basis der signifikanten Ergebnisse bei diesen Endpunkten als „Hinweis“ eingestuft werden.

Für die Gesamtschau des Zusatznutzens sind folgende Ergebnisse relevant:

- **Morbidität:**

- **Beträchtlicher Zusatznutzen** in Hinblick auf die **Vermeidung von Transfusionen**. Die signifikante Reduktion des Risikos für Transfusionen bzw. der Erhalt der Transfusionsfreiheit unter Luspatercept ging in der Studie BEYOND gleichzeitig mit einer deutlichen Steigerung und langfristigen Stabilisierung der Hb-Werte und somit einer anhaltenden Linderung der Anämie einher. Diese Ergebnisse zeigen, dass die Verbesserung der Erythropoese durch Luspatercept den Patient:innen einerseits den langfristigen Erhalt einer Transfusionsunabhängigkeit – und, damit einhergehend, eine Vermeidung von Transfusionsrisiken, Folgekomplikationen sowie weiteren Belastungen – bietet und ihnen andererseits eine Reduktion der Anämie-assoziierten Symptomatik und Verbesserung ihres Befindens ermöglichen kann.
- **Beträchtlicher Zusatznutzen** hinsichtlich der **Symptomatik gemäß PGI-S und PGI-C**. Als selbstberichtete Fragebögen spiegeln die Endpunkte die unmittelbare Erfahrung und Wahrnehmung der Patient:innen wider. Die Ergebnisse zeigen, dass die Therapie mit Luspatercept gegenüber der zVT zu einer sehr deutlichen und patientenrelevanten Reduktion der durch die Patient:innen wahrgenommenen Symptombelastung führt.

- **Lebensqualität:**

- **Beträchtlicher Zusatznutzen** in Bezug auf das **körperliche und psychische Wohlbefinden (SF-36) sowie den Gesamtscore der Fatigue-spezifischen Lebensqualität (FACIT-F)**. Die Ergebnisse zeigen, dass die Behandlung mit Luspatercept gegenüber der zVT mit einer spürbaren Verbesserung in verschiedenen Bereichen der Lebensqualität verbunden ist. Aufgrund der chronischen Natur der Erkrankung ist die gesundheitsbezogene Lebensqualität bei Patient:innen mit NTDT von besonderer Relevanz.

In der Gesamtschau zeigt sich durch Luspatercept eine beträchtliche Verbesserung des therapie-relevanten Nutzens, begründet in signifikanten und spürbaren Behandlungsunterschieden, die einer deutlichen Verbesserung der Morbidität und gesundheitsbezogenen Lebensqualität der Patient:innen entsprechen.

**Insgesamt** ergibt sich für die Behandlung von erwachsenen Patient:innen mit Anämie, die mit einer nicht-transfusionsabhängigen Beta-Thalassämie verbunden ist, auf Basis der Ergebnisse der Studie BEYOND ein **Hinweis auf einen beträchtlichen Zusatznutzen von Luspatercept** gegenüber der 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 4-83: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
Erwachsene Patient:innen mit Anämie, die mit nicht-transfusionsabhängiger Beta-Thalassämie verbunden ist	Beträchtlich

#### 4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte

##### 4.5.1 Begründung für die Vorlage indirekter Vergleiche

Sofern mit dem Dossier indirekte Vergleiche (Abschnitt 4.3.2.1) eingereicht wurden, begründen Sie dies. Begründen Sie dabei auch, warum sich die ausgewählten Studien jeweils für einen indirekten Vergleich gegenüber dem zu bewertenden Arzneimittel und damit für den Nachweis eines Zusatznutzens durch indirekten Vergleich eignen.

Nicht zutreffend.

##### 4.5.2 Begründung für die Vorlage 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 2005<sup>19</sup>, Molenberghs 2010<sup>20</sup>). 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 2006<sup>21</sup>) 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 2006<sup>22</sup>) 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)

<sup>19</sup> Burzykowski T (Ed.): The evaluation of surrogate endpoints. New York: Springer; 2005.

<sup>20</sup> 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.

<sup>21</sup> Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006; 5(3): 173-186.

<sup>22</sup> Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; 25(2): 183-203.

- 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.

#### 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. European Medicines Agency (EMA) (2023): Orphan Maintenance Assessment Report of an orphan medicinal product submitted for type II variation application - Reblozyl (luspatercept) - Treatment of beta-thalassaemia intermedia and major EU/3/14/1300. [Zugriff: 21.03.2023]. URL: [https://www.ema.europa.eu/en/documents/orphan-maintenance-report-post/reblozyl-orphan-maintenance-assessment-report-post-authorisation\\_en.pdf](https://www.ema.europa.eu/en/documents/orphan-maintenance-report-post/reblozyl-orphan-maintenance-assessment-report-post-authorisation_en.pdf).
2. Institut für Qualität und Wirtschaftlichkeit (IQWiG) (2022): Allgemeine Methoden Version 6.1. [Zugriff: 20.02.2023]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
3. Bristol-Myers Squibb Pharma EEIG (2020): Reblozyl® 25 mg / 75 mg Pulver zur Herstellung einer Injektionslösung; Fachinformation. Stand: Februar 2023 [Zugriff: 14.03.2023]. URL: <http://www.fachinfo.de>.
4. Gemeinsamer Bundesausschuss (G-BA) (2022): Niederschrift (finale Fassung) zum Beratungsgespräch gemäß § 8 AM-NutzenV Beratungsanforderung 2021-B-292 Luspatercept zur Behandlung nicht transfusionsabhängiger Anämie aufgrund von Beta-Thalassämie.
5. Sleiman J, Tarhini A, Bou-Fakhredin R, Saliba AN, Cappellini MD, Taher AT (2018): Non-Transfusion-Dependent Thalassemia: An Update on Complications and Management. International Journal of Molecular Sciences; 19(1):182.
6. Musallam KM, Rivella S, Vichinsky E, Rachmilewitz EA (2013): Non-transfusion-dependent thalassemias. Haematologica; 98(6):833-44.
7. Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini M, Del Vecchio G, et al. (2004): Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. Haematologica; 89(10):1187-93.
8. Davis BA, O'Sullivan C, Jarritt PH, Porter JB (2004): Value of sequential monitoring of left ventricular ejection fraction in the management of thalassemia major. Blood; 104(1):263-9.

9. Musallam KM, Cappellini MD, Wood JC, Motta I, Graziadei G, Tamim H, et al. (2011): Elevated liver iron concentration is a marker of increased morbidity in patients with  $\beta$  thalassemia intermedia. *Haematologica*; 96(11):1605-12.
10. Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO) (2022): Beta Thalassämie Leitlinie. [Zugriff: 20.02.2023]. URL: <https://www.onkopedia.com/de/onkopedia/guidelines/beta-thalassaemie/@@guideline/html/index.html>.
11. Gesellschaft für Pädiatrische Onkologie und Hämatologie (GPOH) (2022): S2k-Leitlinie Diagnostik und Therapie der sekundären Eisenüberladung bei Patienten mit angeborenen Anämien. [Zugriff: 20.02.2023]. URL: [https://register.awmf.org/assets/guidelines/025-0291\\_S2k\\_Diagnostik-Therapie-der-sekundaeren-Eisenueberladung-bei-Patienten-mit-angeborenen-Anaemien\\_2022-03\\_1.pdf](https://register.awmf.org/assets/guidelines/025-0291_S2k_Diagnostik-Therapie-der-sekundaeren-Eisenueberladung-bei-Patienten-mit-angeborenen-Anaemien_2022-03_1.pdf).
12. Shah FT, Sayani F, Trompeter S, Drasar E, Piga A (2019): Challenges of blood transfusions in  $\beta$ -thalassemia. *Blood Reviews*; 37:100588.
13. Taher A, Viprakasit V, Cappellini MD, Sutcharitchan P, Ward R, Mahmoud D, et al. (2019): Development of a patient-reported outcomes symptom measure for patients with nontransfusion-dependent thalassemia (NTDT-PRO<sup>®</sup>). *American journal of hematology*; 94(2):171-6.
14. Taher A, Cappellini MD, Viprakasit V, Sutcharitchan P, Mahmoud D, Laadem A, et al. (2019): Validation of a patient-reported outcomes symptom measure for patients with nontransfusion-dependent thalassemia (NTDT-PRO<sup>®</sup>). *American journal of hematology*; 94(2):177-83.
15. Gemeinsamer Bundesausschuss (G-BA) (2020): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo – Wirkstoff: Givosiran. [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/92-975-3680/2020-04-15\\_Nutzenbewertung-G-BA\\_Givosiran\\_D-536.pdf](https://www.g-ba.de/downloads/92-975-3680/2020-04-15_Nutzenbewertung-G-BA_Givosiran_D-536.pdf).
16. Gemeinsamer Bundesausschuss (G-BA) (2019): Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Nintedanib (Neubewertung eines Orphan-Drugs nach Überschreitung der 50 Mio. Euro Grenze) [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/40-268-6058/2019-10-17\\_AM-RL-XII\\_Nintedanib\\_D-450\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-6058/2019-10-17_AM-RL-XII_Nintedanib_D-450_TrG.pdf).
17. Gemeinsamer Bundesausschuss (G-BA) (2021): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo – Wirkstoff: Fedratinib. [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/92-975-4518/2021-03-15\\_Nutzenbewertung-G-BA\\_Fedratinib\\_D-650.pdf](https://www.g-ba.de/downloads/92-975-4518/2021-03-15_Nutzenbewertung-G-BA_Fedratinib_D-650.pdf).
18. Gemeinsamer Bundesausschuss (G-BA) (2020): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo – Wirkstoff: Burosumab. [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/92-975-3331/2019-10-01\\_Nutzenbewertung-G-BA\\_Burosumab\\_D-492.pdf](https://www.g-ba.de/downloads/92-975-3331/2019-10-01_Nutzenbewertung-G-BA_Burosumab_D-492.pdf).
19. Gemeinsamer Bundesausschuss (G-BA) (2019): 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 - Inotersen. Vom 22. März 2019. [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/39-261-3717/2019-03-22\\_AM-RL-XII\\_Inotersen\\_D-381\\_BAnz.pdf](https://www.g-ba.de/downloads/39-261-3717/2019-03-22_AM-RL-XII_Inotersen_D-381_BAnz.pdf).
20. Gemeinsamer Bundesausschuss (G-BA) (2019): 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 - Inotersen. Vom 22. März 2019. [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/40-268-5646/2019-03-22\\_AM-RL-XII\\_Inotersen\\_D-381\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-5646/2019-03-22_AM-RL-XII_Inotersen_D-381_TrG.pdf).
  21. McHorney CA, Ware JE, Jr., Raczek AE (1993): The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*; 31(3):247-63.
  22. Ware JE, Jr., Sherbourne CD (1992): The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*; 30(6):473-83.
  23. Sobota A, Yamashita R, Xu Y, Trachtenberg F, Kohlbry P, Kleinert DA, et al. (2011): Quality of life in thalassemia: a comparison of SF-36 results from the thalassemia longitudinal cohort to reported literature and the US norms. *Am J Hematol*; 86(1):92-5.
  24. Haghpanah S, Nasirabadi S, Ghaffarparand F, Karami R, Mahmoodi M, Parand S, et al. (2013): Quality of life among Iranian patients with beta-thalassemia major using the SF-36 questionnaire. *Sao Paulo Medical Journal*; 131:166-72.
  25. Porter J, Bowden DK, Economou M, Troncy J, Ganser A, Habr D, et al. (2012): Health-Related Quality of Life, Treatment Satisfaction, Adherence and Persistence in  $\beta$ -Thalassemia and Myelodysplastic Syndrome Patients with Iron Overload Receiving Deferasirox: Results from the EPIC Clinical Trial. *Anemia*; 2012:10.
  26. Javanbakht M, Keshtkaran A, Shabaninejad H, Karami H, Zakerinia M, Delavari S (2015): Comparison of Blood Transfusion Plus Chelation Therapy and Bone Marrow Transplantation in Patients with  $\beta$ -Thalassemia: Application of SF-36, EQ-5D, and Visual Analogue Scale Measures. *Int J Health Policy Manag*; 4(11):733-40.
  27. Institut für Qualität und Wirtschaftlichkeit (IQWiG) (2021): Filgotinib (rheumatoide Arthritis) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. [Zugriff: 20.02.2023]. URL: [https://www.iqwig.de/download/a20-90\\_filgotinib\\_nutzenbewertung-35a-sgb-v\\_v1-0.pdf](https://www.iqwig.de/download/a20-90_filgotinib_nutzenbewertung-35a-sgb-v_v1-0.pdf).
  28. Gemeinsamer Bundesausschuss (G-BA) (2020): Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Upadacitinib (Rheumatoide Arthritis). [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/40-268-6731/2020-07-16\\_AM-RL-XII\\_Upadacitinib\\_D-509\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-6731/2020-07-16_AM-RL-XII_Upadacitinib_D-509_TrG.pdf).
  29. Gemeinsamer Bundesausschuss (G-BA) (2020): Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Ravulizumab. [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/40-268-6320/2020-02-06\\_AM-RL-XII\\_Ravulizumab\\_D-463\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-6320/2020-02-06_AM-RL-XII_Ravulizumab_D-463_TrG.pdf).
  30. Gemeinsamer Bundesausschuss (G-BA) (2017): 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 – Ibrutinib (neues Anwendungsgebiet: chronische

- lymphatische Leukämie; in Kombination mit Bendamustin und Rituximab). [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/40-268-4246/2017-03-16\\_AM-RL-XII\\_Ibrutinib\\_nAWG-Laeukemie-D-262\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-4246/2017-03-16_AM-RL-XII_Ibrutinib_nAWG-Laeukemie-D-262_TrG.pdf).
31. Gemeinsamer Bundesausschuss (G-BA) (2021): Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Filgotinib (Rheumatoide Arthritis). [Zugriff: 20.02.2023]. URL: [https://www.g-ba.de/downloads/40-268-7451/2021-04-15\\_AM-RL-XII\\_Filgotinib\\_D-590\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-7451/2021-04-15_AM-RL-XII_Filgotinib_D-590_TrG.pdf).
  32. Mayo Clinic (2021): Thalassemia [Zugriff: 20.02.2023]. URL: <https://www.mayoclinic.org/diseases-conditions/thalassemia/symptoms-causes/syc-20354995>.
  33. Taher AT, Musallam KM, Cappellini MD (2021):  $\beta$ -Thalasseмии. New England Journal of Medicine; 384(8):727-43.
  34. U.S. Department of Health and Human Services (2009): Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. [Zugriff: 20.02.2023]. URL: [https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE\\_4.03/CTCAE\\_4.03\\_2010-06-14\\_QuickReference\\_8.5x11.pdf](https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf).
  35. Taher A, Vichinsky E, Musallam K, Cappellini MD, Viprakasit V (2017): Guidelines for the Management of non transfusion dependent Thalassaemia (NTDT), 2<sup>nd</sup> Edition. [Zugriff: 20.02.2023]. URL: <https://thalassaemia.org.cy/wp-content/uploads/2017/10/NTDT-final-combined-1.pdf>.
  36. Taher AT, Musallam KM, Saliba AN, Graziadei G, Cappellini MD (2015): Hemoglobin level and morbidity in non-transfusion-dependent thalassemia. Blood cells, molecules & diseases; 55(2):108-9.
  37. Novartis Europharm Limited (2006): EXJADE® Filmtabletten; Fachinformation. Stand: August 2022 [Zugriff: 14.03.2023]. URL: <http://www.fachinfo.de>.
  38. Chiesi Farmaceutici S.p.A. (1999): Ferriprox 500 mg/1000 mg Filmtabletten; Fachinformation. Stand: September 2022 [Zugriff: 14.03.2023]. URL: <http://www.fachinfo.de>.
  39. Celgene (2017): ACE-536-B-THAL-002 - A Study to Determine the Efficacy and Safety of Luspatercept in Adults With Non Transfusion Dependent Beta ( $\beta$ )-Thalassemia (BEYOND) (NCT03342404). Stand des Eintrags: 19.12.2022. [Zugriff: 12.02.2023]. URL: <https://clinicaltrials.gov/ct2/show/record/NCT03342404?view=record>
  40. Celgene (2017): ACE-536-B-THAL-002 - A Study to Determine the Efficacy and Safety of Luspatercept in Adults With Non Transfusion Dependent Beta ( $\beta$ )-Thalassemia BEYOND - WHO ICTRP (NCT03342404). Stand des Eintrags: 09.01.2023. [Zugriff: 15.02.2023]. URL: <https://trialsearch.who.int/Trial2.aspx?TrialID=NCT03342404>
  41. Taher AT, Cappellini MD, Kattamis A, Voskaridou E, Perrotta S, Piga AG, et al. (2022): Luspatercept for the treatment of anaemia in non-transfusion-dependent  $\beta$ -thalassaemia (BEYOND): a phase 2, randomised, double-blind, multicentre, placebo-controlled trial. The Lancet Haematology; 9(10):e733-e44.
  42. Celgene (2021): Luspatercept (Ace-536) Clinical Study Report Ace-536-B-Thal-002. A Phase 2, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study To Determine The Efficacy And Safety Of Luspatercept (Ace-536) Versus Placebo In

- Adults With Non-Transfusion Dependent Beta (B)-Thalassemia (THE BEYOND™ STUDY).
43. Musallam KM, Cappellini MD, Daar S, Taher AT (2022): Morbidity-free survival and hemoglobin level in non-transfusion-dependent  $\beta$ -thalassemia: a 10-year cohort study. *Annals of Hematology*; 101(1):203-4.
  44. Musallam KM, Cappellini MD, Taher AT (2021): Variations in hemoglobin level and morbidity burden in non-transfusion-dependent  $\beta$ -thalassemia. *Annals of Hematology*; 100(7):1903-5.
  45. Cappellini MD, Viprakasit V, Taher AT, Georgiev P, Kuo KH, Coates T, et al. (2020): A phase 3 trial of luspatercept in patients with transfusion-dependent  $\beta$ -thalassemia. *New England Journal of Medicine*; 382(13):1219-31.
  46. Gesellschaft für Pädiatrische Onkologie und Hämatologie (GPOH) (2020): AWMF-Leitlinie 025/016 „Sichelzellerkrankheit“. [Zugriff: 14.03.2023]. URL: [https://register.awmf.org/assets/guidelines/025-016l\\_S2k\\_Sichelzellerkrankheit\\_2020-12.pdf](https://register.awmf.org/assets/guidelines/025-016l_S2k_Sichelzellerkrankheit_2020-12.pdf).
  47. Cario H, Stahnke K, Kohne E (1999): Beta-Thalassämie in Deutschland. Ergebnisse der kooperativen Beta-Thalassämie-Studie. (beta-Thalassämie in Deutschland. Ergebnisse der kooperativen beta-Thalassämie-Studie.). *Klin Padiatr*; 211(6):431-7.
  48. Cario H, Stahnke K, Sander S, Kohne E (2000): Epidemiological situation and treatment of patients with thalassemia major in Germany: results of the German multicenter beta-thalassemia study. *Ann Hematol*; 79(1):7-12.
  49. Vetter B, Schwarz C, Kohne E, Kulozik AE (1997): Beta-thalassaemia in the immigrant and non-immigrant German populations. *Br J Haematol*; 97(2):266-72.
  50. Kohne E, Kleihauer E (2010): Hemoglobinopathies: a longitudinal study over four decades. *Dtsch Arztebl Int*; 107(5):65-71.
  51. Institut für angewandte Gesundheitsforschung Berlin (InGef) (2022): Studienbericht Epidemiologie und Charakteristika von Patienten mit Beta-Thalassämie in Deutschland - Zusatzanalyse.
  52. Wong SS, Wilczynski NL, Haynes RB (2006): 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 : JMLA*; 94(4):451-5.

**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):

<b>Datenbankname</b>	EMBASE	
<b>Suchoberfläche</b>	Ovid	
<b>Datum der Suche</b>	07.11.2016	
<b>Zeitsegment</b>	1974 to 2016 November 04	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle <sup>23</sup> ] – Strategy minimizing difference between sensitivity and specificity	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
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

<sup>23</sup> 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****Suche nach RCT mit dem zu bewertenden Arzneimittel in MEDLINE®**

<b>Datenbankname</b>	MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily	
<b>Suchoberfläche</b>	Ovid	
<b>Datum der Suche</b>	02.02.2023	
<b>Zeitsegment</b>	1946 bis 01.02.2023	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 [52]	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	exp luspatercept/ OR luspatercept*.mp.	151
2	Reblozyl*.mp.	11
3	(ACE-536 OR ACE536).mp.	12
4	1 OR 2 OR 3	156
5	exp thalassemia/ OR thalass*.mp.	33310
6	randomized controlled trial.pt. OR randomized.mp. OR placebo.mp.	1070819
7	4 AND 5 AND 6	9

**Suche nach RCT mit dem zu bewertenden Arzneimittel in EMBASE**

<b>Datenbankname</b>	EMBASE	
<b>Suchoberfläche</b>	Ovid	
<b>Datum der Suche</b>	02.02.2023	
<b>Zeitsegment</b>	1974 bis 02.02.2023	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 [52]	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	exp luspatercept/ OR luspatercept*.mp.	544
2	Reblozyl*.mp.	24
3	(ACE-536 OR ACE536).mp.	121
4	1 OR 2 OR 3	568
5	exp thalassemia/ OR thalass*.mp.	44041
6	random*.tw. OR placebo*.mp. OR double-blind*.tw.	2175487
7	4 AND 5 AND 6	51

**Suche nach RCT mit dem zu bewertenden Arzneimittel in der Cochrane Library**

<b>Datenbankname</b>	The Cochrane Library	
<b>Suchoberfläche</b>	The Cochrane Library ( <a href="https://www.cochranelibrary.com/advanced-search">https://www.cochranelibrary.com/advanced-search</a> )	
<b>Datum der Suche</b>	02.02.2023	
<b>Zeitsegment</b>	Keine zeitliche Einschränkung	
<b>Suchfilter</b>	Nur klinische Studien [Trials] in 9	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	luspatercept*	103
2	Reblozyl*	1
3	ACE-536 OR ACE536	22
4	#1 OR #2 OR #3	104
5	MeSH term [Thalassemia] explode all trees	453
6	thalass*	1559
7	#5 OR #6	1562
8	#4 AND #7	43
9	davon Trials	43

**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), 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:

<b>Studienregister/ Studienergebnisdatenbank</b>	International Clinical Trials Registry Platform Search Portal
<b>Internetadresse</b>	<a href="http://apps.who.int/trialsearch/">http://apps.who.int/trialsearch/</a>
<b>Datum der Suche</b>	07.11.2016
<b>Eingabeoberfläche</b>	Standard Search
<b>Suchstrategie</b>	linagliptin OR BI 1356
<b>Treffer</b>	169

**Anhang 4-B1: Suche nach RCT mit dem zu bewertenden Arzneimittel****Suche nach RCT mit dem zu bewertenden Arzneimittel in Clinicaltrials.gov**

<b>Studienregister</b>	clinicaltrials.gov
<b>Internetadresse</b>	<a href="http://www.clinicaltrials.gov">http://www.clinicaltrials.gov</a>
<b>Datum der Suche</b>	02.02.2023
<b>Eingabeoberfläche</b>	Advanced Search
<b>Suchstrategie</b>	luspatercept OR Reblozyl OR ACE-536 OR ACE536 [Other Terms] AND thalassemia [Condition/disease]
<b>Treffer</b>	11

**Suche nach RCT mit dem zu bewertenden Arzneimittel im EU Clinical Trials Register (EU-CTR)**

<b>Studienregister</b>	EU Clinical Trials Register
<b>Internetadresse</b>	<a href="https://www.clinicaltrialsregister.eu/ctr-search/search">https://www.clinicaltrialsregister.eu/ctr-search/search</a>
<b>Datum der Suche</b>	02.02.2023
<b>Eingabeoberfläche</b>	Basic Search
<b>Suchstrategie</b>	(luspatercept* OR Reblozyl OR ACE-536 OR ACE536) AND (thalass*)
<b>Treffer</b>	7

**Suche nach RCT mit dem zu bewertenden Arzneimittel im ICTRP Search Portal der WHO**

<b>Studienregister</b>	ICTRP Search Portal der WHO
<b>Internetadresse</b>	<a href="http://apps.who.int/trialsearch/">http://apps.who.int/trialsearch/</a>
<b>Datum der Suche</b>	02.02.2023
<b>Eingabeoberfläche</b>	Basic Search
<b>Suchstrategie</b>	(luspatercept* OR Reblozyl OR ACE-536 OR ACE536) AND (thalass*)
<b>Treffer</b>	37 Einträge zu 19 Studien

**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

Tabelle 4-84 (Anhang): Liste der im Volltext gesichteten und ausgeschlossenen Dokumente der bibliografischen Literaturrecherche – Suche nach RCT

Nummer	Im Volltext gesichtete, ausgeschlossene Studien	Ausschlussgrund
1	Chen, N., Kassir, N., Laadem, A., Giuseppi, A.C., Shetty, J., Maxwell, S.E., Sriraman, P., Ritland, S., Linde, P.G., Budda, B., Reynolds, J.G., Zhou, S. and Palmisano, M. (2021), Population Pharmacokinetics and Exposure-Response Relationship of Luspatercept, an Erythroid Maturation Agent, in Anemic Patients With $\beta$ -Thalassemia. <i>The Journal of Clinical Pharmacology</i> , 61: 52-63. <a href="https://doi.org/10.1002/jcph.1696">https://doi.org/10.1002/jcph.1696</a>	A1

#### 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

Nr.	Referenz	Ausschlussgrund
<b>Clinicaltrials.gov</b>		
1	Celgene (2017): ACE-536-B-THAL-002 U1111-1202-7068 (Registry Identifier: WHO) 2015-003225-33 (EudraCT Number) - A Phase 2, Double-Blind, Randomized, Placebo-Controlled Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) Versus Placebo in Adults With Non-Transfusion Dependent Beta-Thalassemia (The BEYOND Study) - ClinicalTrials.gov (NCT03342404). Stand des Eintrags: 19.12.2022. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT03342404">https://ClinicalTrials.gov/ct2/show/record/NCT03342404</a>	Kein Ausschlussgrund /eingeschlossen
2	Celgene (2019): ACE-536-B-THAL-004 U1111-1241-4168 (Other Identifier: WHO) 2019-000208-13 (EudraCT Number) - A Phase 2a Study to Evaluate the Safety and Pharmacokinetics of Luspatercept (ACE-536) in Pediatric Participants Who Require Regular Red Blood Cell Transfusions Due to Beta-Thalassemia - ClinicalTrials.gov (NCT04143724). Stand des Eintrags: 01.02.2023. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT04143724">https://ClinicalTrials.gov/ct2/show/record/NCT04143724</a>	A7
3	Bristol-Myers Squibb (2022): CA056-001 - A Phase 2, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Evaluate the Efficacy, Safety and Pharmacokinetics of Luspatercept (ACE-536) in Chinese Adult Subjects Who Require Regular Red Blood Cell Transfusions Due to Beta-Thalassemia - ClinicalTrials.gov (NCT05567458). Stand des Eintrags: 17.01.2023. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT05567458">https://ClinicalTrials.gov/ct2/show/record/NCT05567458</a>	A7

Nr.	Referenz	Ausschlussgrund
4	Acceleron Pharma Inc. (a wholly owned subsidiary of Merck Sharp and Dohme, a subsidiary of Merck & Co., Inc.) (2014): A536-06 - An Open-Label Extension Study to Evaluate the Long-Term Effects of ACE-536 in Patients With Beta-Thalassemia Previously Enrolled in Study A536-04 - ClinicalTrials.gov (NCT02268409). Stand des Eintrags: 19.05.2021. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT02268409">https://ClinicalTrials.gov/ct2/show/record/NCT02268409</a>	A5
5	Acceleron Pharma Inc. (a wholly owned subsidiary of Merck Sharp and Dohme, a subsidiary of Merck & Co., Inc.) (2012): A536-04 2012-002499-15 (EudraCT Number) - A Phase 2, Open-Label, Ascending Dose Study to Evaluate the Effects of ACE-536 in Patients With Beta-Thalassemia Intermedia - ClinicalTrials.gov (NCT01749540). Stand des Eintrags: 14.12.2016. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT01749540">https://ClinicalTrials.gov/ct2/show/record/NCT01749540</a>	A5
6	Sun Yat-sen University (2022): SYSKY-2022-088-02 - The Safety and Efficiency of Luspatercept in Chinese Adults With Transfusion Dependent Beta-thalassemia: a Real-world Study - ClinicalTrials.gov (NCT05462548). Stand des Eintrags: 18.07.2022. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT05462548">https://ClinicalTrials.gov/ct2/show/record/NCT05462548</a>	A1
7	Celgene (2015): ACE-536-B-THAL-001 - A Phase 3, Double-Blind, Placebo Controlled Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) in Adults With Transfusion Dependent Beta (B)-Thalassemia - ClinicalTrials.gov (NCT02604433). Stand des Eintrags: 25.01.2022. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT02604433">https://ClinicalTrials.gov/ct2/show/record/NCT02604433</a>	A1
8	Bristol-Myers Squibb (2022): CA056-015 2021-004928-15 (EudraCT Number) - A Phase 2, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (BMS-986346/ACE-536) for the Treatment of Anemia in Adults With Alpha ( $\alpha$ )-Thalassemia - ClinicalTrials.gov (NCT05664737). Stand des Eintrags: 04.01.2023. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT05664737">https://ClinicalTrials.gov/ct2/show/record/NCT05664737</a>	A7
9	Celgene (2021): ACE-536-MDS-005 - A Study to Evaluate the Effectiveness of the Additional Risk Minimisation Measures (aRMMs) for REBLOZYL Among Healthcare Professionals (HCPs) in the European Economic Area (EEA) - ClinicalTrials.gov (NCT04973280). Stand des Eintrags: 06.09.2022. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT04973280">https://ClinicalTrials.gov/ct2/show/record/NCT04973280</a>	A7

Nr.	Referenz	Ausschlussgrund
10	Celgene (2015): ACE-536-B-THAL-003 - Burden of Beta-Thalassemia - Quality of Life and Health Care Resource Utilization- A Prospective Observational Study - ClinicalTrials.gov (NCT02626689). Stand des Eintrags: 07.04.2017. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT02626689">https://ClinicalTrials.gov/ct2/show/record/NCT02626689</a>	A7
11	Celgene (2019): ACE-536-LTFU-001 U1111-1235-8123 (Registry Identifier: WHO) 2018-002915-93 (EudraCT Number) - A Phase 3b, Open-label, Single-arm, Rollover Study to Evaluate Long-term Safety in Subjects Who Have Participated in Other Luspatercept (ACE-536) Clinical Trials - ClinicalTrials.gov (NCT04064060). Stand des Eintrags: 01.02.2023. [Zugriff: 02.02.2023]. URL: <a href="https://ClinicalTrials.gov/ct2/show/record/NCT04064060">https://ClinicalTrials.gov/ct2/show/record/NCT04064060</a>	A7
<b>EU Clinical Trials Register (EU-CTR)</b>		
12	Celgene Corporation (2019): ACE-536-LTFU-001 - A phase 3B, open-label, single-arm, rollover study to evaluate long-term safety in subjects who have participated in other Luspatercept (ACE-536) clinical trials. - EU-CTR (2018-002915-93). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2018-002915-93">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2018-002915-93</a>	A7
13	Celgene Corporation (2022): CA056-015 - A Phase 2, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (BMS-986346/ACE-536) for the Treatment of Anemia in Adults with Alpha ( $\alpha$ )-thalassemia - EU-CTR (2021-004928-15). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2021-004928-15">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2021-004928-15</a>	A7
14	Celgene Corporation (2019): ACE-536-B-THAL-004 - A phase 2a study to evaluate the safety and pharmacokinetics of Luspatercept (ACE-536) in paediatric subjects who require regular red blood cell transfusions due to beta thalassemia - EU-CTR (2019-000208-13). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2019-000208-13">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2019-000208-13</a>	A1
15	Celgene Corporation (2015): ACE-536-B-THAL-001 - A Phase 3, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) Versus Placebo in Adults Who Require Regular Red Blood Cell Transfusions Due to Beta-Thalassemia - EU-CTR (2015-003224-31). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2015-003224-31">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2015-003224-31</a>	A1

Nr.	Referenz	Ausschlussgrund
16	Celgene Corporation (2017): ACE-536-B-THAL-002 - A Phase 2, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) versus Placebo in Adults with Non Transfusion Dependent Beta-Thalassemia - EU-CTR (2015-003225-33). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2015-003225-33">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2015-003225-33</a>	A7
17	ACCELERON PHARMA INC (2012): A536-04 - A Phase 2, Open-Label, Ascending Dose Study to Evaluate the Effects of ACE-536 in Patients with $\beta$ -Thalassemia Intermedia - EU-CTR (2012-002499-15). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012-002499-15">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012-002499-15</a>	A5
18	Acceleron Pharma, Inc. (2014): A536-06 - An Open-Label Extension Study to Evaluate the Long-Term Effects of ACE-536 in Patients with Beta-Thalassemia Previously Enrolled in Study A536-04 - EU-CTR (2014-001281-94). [Zugriff: 02.02.2023]. URL: <a href="https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2014-001281-94">https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2014-001281-94</a>	A5
<b>ICTRP Search Register</b>		
19	Bristol-Myers Squibb (2022): CA056-015 2021-004928-15 - A Phase 2, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (BMS-986346/ACE-536) for the Treatment of Anemia in Adults With Alpha (a)-Thalassemia - WHO ICTRP (NCT05664737). Stand des Eintrags: 01.01.2023. [Zugriff: 02.02.2023]. URL: <a href="https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05664737">https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05664737</a>	A7
20	Bristol-Myers Squibb (2022): CA056-001 - A Phase 2, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Evaluate the Efficacy, Safety and Pharmacokinetics of Luspatercept (ACE-536) in Chinese Adult Subjects Who Require Regular Red Blood Cell Transfusions Due to Beta ( $\beta$ )-Thalassemia - WHO ICTRP (NCT05567458). Stand des Eintrags: 02.01.2023. [Zugriff: 02.02.2023]. URL: <a href="https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05567458">https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05567458</a>	A7
21	Sun Yat-sen University (2022): - The Safety and Efficiency of Luspatercept in Chinese Adults With Transfusion Dependent Beta-thalassemia: a Real-world Study - WHO ICTRP (NCT05462548). Stand des Eintrags: 02.07.2022. [Zugriff: 02.02.2023]. URL: <a href="https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05462548">https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05462548</a>	A1

Nr.	Referenz	Ausschlussgrund
22	Celgene Corporation (2022): IND 112562 - A Phase 3B, open-label, single-arm, rollover study to evaluate long-term safety in subjects who have participated in other luspatercept (ACE-536) clinical trials - WHO ICTRP (LBCTR2019100218). Stand des Eintrags: 09.01.2023. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=LBCTR2019100218">https://trialssearch.who.int/Trial2.aspx?TrialID=LBCTR2019100218</a>	A7
23	Celgene (2021): ACE-536-MDS-005 - A Study to Evaluate the Effectiveness of the Additional Risk Minimisation Measures (aRMMs) for REBLOZYL Among Healthcare Professionals (HCPs) in the European Economic Area (EEA) - WHO ICTRP (NCT04973280). Stand des Eintrags: 01.09.2022. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=NCT04973280">https://trialssearch.who.int/Trial2.aspx?TrialID=NCT04973280</a>	A7
24	Celgene Corporation (2019): ACE-536-LTFU-001 2018-002915-93-SE - WHO ICTRP (2018-002915-93). Stand des Eintrags: 03.04.2019. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-GR">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-GR</a>	A5
25	Celgene Corporation (2019): 2018-002915-93-SE ACE-536-LTFU-001 - WHO ICTRP (2018-002915-93). Stand des Eintrags: 03.06.2019. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-FR">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-FR</a>	A5
26	Celgene Corporation (2019): 2018-002915-93-DE ACE-536-LTFU-001 - A phase 3B, open-label, single-arm, rollover study to evaluate long-term safety in subjects who have participated in other Luspatercept (ACE-536) clinical trials. - WHO ICTRP (2018-002915-93). Stand des Eintrags: 01.01.2022. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-NL">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-NL</a>	A7
27	Celgene Corporation (2019): ACE-536-LTFU-001 - A phase 3B, open-label, single-arm, rollover study to evaluate long-term safety in subjects who have participated in other Luspatercept (ACE-536) clinical trials. - WHO ICTRP (2018-002915-93). Stand des Eintrags: 02.11.2021. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-DE">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2018-002915-93-DE</a>	A5

Nr.	Referenz	Ausschlussgrund
28	Celgene (2017): ACE-536-B-THAL-002 2015-003225-33 U1111-1202-7068 - A Phase 2, Double-Blind, Randomized, Placebo-Controlled Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) Versus Placebo in Adults With Non-Transfusion Dependent Beta ( $\beta$ )-Thalassemia (The BEYOND Study) - WHO ICTRP (NCT03342404). Stand des Eintrags: 09.01.2023. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=NCT03342404">https://trialssearch.who.int/Trial2.aspx?TrialID=NCT03342404</a>	Kein Ausschlussgrund /eingeschlossen
29	Celgene Corporation (2017): ACE-536-B-THAL-002 2015-003225-33-GB - A Phase 2, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) versus Placebo in Adults with Non Transfusion Dependent Beta ( $\beta$ )-Thalassemia - The BEYOND Study - WHO ICTRP (2015-003225-33). Stand des Eintrags: 01.09.2020. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2015-003225-33-GR">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2015-003225-33-GR</a>	A7
30	Celgene Corporation (2017): ACE-536-B-THAL-002 - A Phase 2, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) versus Placebo in Adults with Non Transfusion Dependent Beta ( $\beta$ )-Thalassemia - The BEYOND Study - WHO ICTRP (2015-003225-33). Stand des Eintrags: 07.09.2020. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2015-003225-33-GB">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2015-003225-33-GB</a>	A7
31	Celgene Corporation (2015): ACE-536-B-THAL-001 - A Phase 3, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) Versus Placebo in Adults Who Require Regular Red Blood Cell Transfusions Due to Beta-Thalassemia - The Believe Trial - WHO ICTRP (2015-003224-31). Stand des Eintrags: 09.11.2020. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2015-003224-31-GB">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2015-003224-31-GB</a>	A1
32	Celgene (2015): ACE-536-B-THAL-003 - Burden of Beta-Thalassemia - Quality of Life and Health Care Resource Utilization- A Prospective Observational Study - WHO ICTRP (NCT02626689). Stand des Eintrags: 01.12.2017. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02626689">https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02626689</a>	A7

Nr.	Referenz	Ausschlussgrund
33	Celgene (2015): ACE-536-B-THAL-001 - A Phase 3, Double-Blind, Placebo Controlled Multicenter Study to Determine the Efficacy and Safety of Luspatercept (ACE-536) in Adults With Transfusion Dependent Beta (β)-Thalassemia - WHO ICTRP (NCT02604433). Stand des Eintrags: 03.01.2022. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02604433">https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02604433</a>	A1
34	Acceleron Pharma Inc. (2014): A536-06 - An Open-Label Extension Study to Evaluate the Long-Term Effects of ACE-536 in Patients With β-Thalassemia Previously Enrolled in Study A536-04 - WHO ICTRP (NCT02268409). Stand des Eintrags: 07.06.2021. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02268409">https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02268409</a>	A5
35	Acceleron Pharma, Inc. (2014): 2014-001281-94-IT A536-06 - An Open-Label Extension Study to Evaluate the Long-Term Effects of ACE-536 in Patients with β-Thalassemia Previously Enrolled in Study A536-04 - WHO ICTRP (2014-001281-94). Stand des Eintrags: 02.05.2020. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2014-001281-94-GR">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2014-001281-94-GR</a>	A5
36	Acceleron Pharma, Inc. (2014): A536-06 - An Open-Label Extension Study to Evaluate the Long-Term Effects of ACE-536 in Patients with β-Thalassemia Previously Enrolled in Study A536-04 - WHO ICTRP (2014-001281-94). Stand des Eintrags: 01.07.2015. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2014-001281-94-IT">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2014-001281-94-IT</a>	A5
37	ACCELERON PHARMA INC (2012): A536-04 - A Phase 2, Open-Label, Ascending Dose Study to Evaluate the Effects of ACE-536 in Patients with β-Thalassemia Intermedia - WHO ICTRP (2012-002499-15). Stand des Eintrags: 01.04.2016. [Zugriff: 02.02.2023]. URL: <a href="https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2012-002499-15-IT">https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2012-002499-15-IT</a>	A5

**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 4-85 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 4-85 inklusive eines Flow-Charts für den Patientenfluss.*

Tabelle 4-85 (Anhang): Studiendesign und -methodik für Studie BEYOND

Item <sup>a</sup>	Charakteristikum	Studieninformation
<b>Studienziel</b>		
<b>2 b</b>	Genaue Ziele, Fragestellung und Hypothesen	<p>Zielsetzung und Fragestellung:</p> <ul style="list-style-type: none"> <li>• Primäres Ziel: Untersuchung der Wirksamkeit und Sicherheit von Luspatercept (ACE-536) im Vergleich zu Placebo zur Behandlung der Anämie bei nicht-transfusionsabhängigen Beta-Thalassämie-Patient:innen, gemessen anhand der mittleren Hämoglobin-Konzentration in Abwesenheit von Transfusionen über einen Zeitraum von 12 Wochen, zwischen Woche 13 und 24, im Vergleich zur Baseline</li> <li>• Sekundäre Ziele: <ul style="list-style-type: none"> <li>○ Untersuchung des Effekts von Luspatercept gegenüber Placebo auf Anämie-bezogene Symptome bei Beta-Thalassämie-Patient:innen, gemessen anhand des NTDT-PRO über 2 kontinuierliche 12-Wochen Intervalle (Woche 13 bis Woche 24 und Woche 37 bis Woche 48), im Vergleich zur Baseline</li> <li>○ Untersuchung des Effekts von Luspatercept gegenüber Placebo auf die funktionelle und gesundheitsbezogene QoL, erhoben anhand des SF-36 und FACIT-F Fragebogens</li> <li>○ Untersuchung der Langzeiteffekte von Luspatercept gegenüber Placebo bei Anämie, gemessen anhand der mittleren Hämoglobin-Konzentration in Abwesenheit von Transfusionen über einen kontinuierlichen Zeitraum eines 12-Wochen Intervalls von Woche 37 bis Woche 48, im Vergleich zur Baseline</li> <li>○ Untersuchung des Effekts von Luspatercept gegenüber Placebo auf die Eisenüberladung, gemessen anhand der LIC und der täglichen Eisenchelator dosis</li> <li>○ Untersuchung des Effekts von Luspatercept gegenüber Placebo auf die Eisenüberladung, gemessen anhand des Serum Ferritin</li> <li>○ Untersuchung der Dauer des erythroiden Ansprechens</li> <li>○ Untersuchung des Effekts von Luspatercept gegenüber Placebo auf die körperliche Leistungsfähigkeit, gemessen anhand des 6-Minuten Gehtests (6MWT)</li> <li>○ Untersuchung der Sicherheit und Verträglichkeit von Luspatercept, inklusive Immunogenität</li> <li>○ Untersuchung der Populationspharmakokinetik (PK) von Luspatercept bei Patient:innen mit Beta-Thalassämie</li> </ul> </li> <li>• Explorative Ziele: <ul style="list-style-type: none"> <li>○ Untersuchung der Effekte von Luspatercept gegenüber Placebo auf die extramedulläre</li> </ul> </li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Blutbildung (EMH), Knochendichte, Splenomegalie, pulmonale Hypertonie und Beinulcera, falls vorhanden</p> <ul style="list-style-type: none"> <li>○ Untersuchung des Effekts von Luspatercept gegen Placebo auf die Schwere der Beta-Thalassämie anhand des NTDT Severity Score Systems</li> <li>○ Untersuchung des Effekts von Luspatercept gegen Placebo auf die Schwere der Beta-Thalassämie anhand der Parameter zum morbiditätsfreien Überleben</li> <li>○ Untersuchung des Zusammenhangs zwischen den Ausgangswerten und der Veränderung des Serum Growth Differentiation Factor 11 (GDF11) und anderer verwandter Biomarker bzgl. der Behandlung mit Luspatercept</li> <li>○ Untersuchung des Effekts von Luspatercept auf das fetale Hämoglobin (HbF)</li> <li>○ Untersuchung des Effekts von Luspatercept auf die Nutzung von Gesundheitsressourcen (HRU)</li> </ul> <p>Hypothese: Die Behandlung mit Luspatercept führt im Vergleich zu Placebo zu einer höheren Wahrscheinlichkeit, über einen Zeitraum von 12 Wochen (Woche 13 bis 24) in Abwesenheit von Transfusionen eine Erhöhung der durchschnittlichen Hämoglobin-Konzentration um 1,0 g/dl zu erreichen.</p>
<b>Methoden</b>		
<b>3</b>	Studiendesign	
<b>3a</b>	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	<p><b>Design:</b> zweiarmig, parallel, Placebo-kontrolliert  <b>Zuteilung der Studienteilnehmer:</b> randomisiert im Verhältnis 2:1 (Luspatercept:Placebo)  <b>Verblindung:</b> doppelblind  <b>Studiendauer:</b> mindestens 48 Wochen pro Patient:in (doppelblinde Behandlungsphase)  <b>Studienorganisation:</b> multizentrisch  <b>Phase:</b> II</p>
<b>3b</b>	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<p><b>Amendment 1</b> (21.12.2018):</p> <ul style="list-style-type: none"> <li>• Dosisintervallverlängerung für eine Verschlechterung der Anämie und Abbruch der Behandlung im Falle bestätigter maligner hämatologischer Veränderungen</li> <li>• Dosisintervallverlängerung bei Grad <math>\geq 3</math> Leukopenie, Neutropenie und/oder Thrombozytopenie und Abbruch der Behandlung im Falle bestätigter maligner hämatologischer Veränderungen</li> </ul> <p><b>Amendment 2</b> (12.06.2020):</p> <ul style="list-style-type: none"> <li>• Die Open-Label Phase nach Entblindung wurde ergänzt, um Placebo-behandelten Patient:innen den Cross-over zur Behandlung mit Luspatercept zu erlauben sowie</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Patient:innen, die bereits von einer Behandlung mit Luspatercept profitierten, für bis zu 15 Monate die Weiterbehandlung zu ermöglichen, bevor ein Wechsel in die Rollover Studie erfolgen sollte</p> <ul style="list-style-type: none"> <li>• Die NTDT-PRO Analysen wurden modifiziert, um den Kenntnisstand der Literatur widerzuspiegeln. So ist kein linearer Zusammenhang zwischen Hämoglobin und patienten-berichteter Fatigue oder HRQoL zu erwarten und die maximale Verbesserung der HRQoL bei Hb-Werten von 11,0 bis 13,0 g/dl zu beobachten (Shasha, 2004; Crawford, 2002).</li> <li>• Hinzufügung des PGI-S als präferierten Ankerpunkt, um im Rahmen eines <i>multiple anchor</i> Ansatzes die von Proband:innen wahrgenommene Schwere der Symptome direkt widerzuspiegeln. Der PGI-S wurde bereits zusammen mit dem NTDT-PRO erhoben, aber explizit zur Beurteilung in das überarbeitete Protokoll hinzugefügt.</li> <li>• Ersetzen des NTDT-PRO Gesamtscores durch den NTDT-PRO Gesamtscore in der Domäne SoB in den sekundären Wirksamkeitsendpunkten (da bislang kein Algorithmus zur Berechnung eines umfassenden Gesamtscores entwickelt wurde).</li> <li>• Die Angabe zur maximalen Dosis pro Gabe wurde aus den Studien-Dosierungs-Richtlinien entfernt, um bei den Dosis-Empfehlungen mit dem zugelassenen Label konsistent zu sein</li> <li>• Ergänzung thromboembolischer Ereignisse als zu untersuchende UE (<i>targeted safety events</i>).</li> <li>• Spezifizierung der drei sekundären Hauptendpunkte sowie der sequentiellen Testprozedur</li> <li>• Die Analyse des Anteils an Proband:innen mit einem Anstieg des durchschnittlichen Scores von <math>\geq 3</math> über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13 bis 24 und Woche 37 bis 48) im Vergleich zur Baseline wurde für die Fatigue-spezifische Skala des FACIT-F (FACIT-Fatigue Skala) spezifiziert</li> <li>• Die Analyse der Änderung des SF-36 zu Woche 24 und Woche 48 gegenüber Baseline wurde für die PCS und MCS spezifiziert</li> <li>• Die Analyse der Anzahl an Proband:innen mit einer Abnahme des durchschnittlichen NTDT-PRO Gesamtscores in der Domäne T/W (Woche 13-24 und Woche 37 bis 48) wurde als geplanter Endpunkt konkretisiert</li> <li>• Definition der HRQoL Populationen und genauere Beschreibung der HRQoL Instrumente</li> <li>• Konkretisierung, dass die primäre Wirksamkeitsanalyse erst stattfinden sollte, wenn alle Patient:innen die 48-wöchige doppelblinde Behandlungsphase (DBTP) abgeschlossen oder abgebrochen hatten und dass zum Ende der Open Label Phase aktualisierte Sicherheits- und Wirksamkeitsanalysen durchgeführt werden sollten.</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p><b>Amendment 3 (27.04.2022):</b></p> <ul style="list-style-type: none"> <li>• Parameter zum morbiditätsfreien Überleben Reduktion auf 7 Parameter, die das "morbiditätsfreie Überleben" definieren, anstelle der 20 Parameter, die im NTD Severity Score-System enthalten sind. Diese 7 Parameter (extramedulläre Hämatopoese (EMH), Lebererkrankung, den Stützapparat betreffend (Skelett), Thrombose, Endokrinopathien, pulmonale Hypertonie und Beinulcera) wurden ausgewählt, weil ihre Beobachtung die Wirksamkeit von Luspatercept besser widerspiegeln dürfte. Ziel dieser Änderung ist es, das Fortschreiten des morbiditätsfreien Überlebens nach Dr. Musallam über einen langfristigen Zeitraum (einschließlich Baseline, Woche 24 und 48 der doppelblinden Behandlungsphase und Woche 144 in der Open-Label-Phase) zu beobachten und zu erfassen.</li> <li>• Dauer der Open-Label-Phase Verlängerung der ursprünglich geplanten Dauer der Open-Label-Phase von 15 Monaten um einige Monate. Ziel dieser Änderung ist es eine Unterbrechung der Behandlung aufgrund dieser zeitlichen Begrenzung zu vermeiden. Dies sollte es allen Probanden (unter Behandlung in der Open-Label-Phase oder in der Posttreatment-Follow-up-Phase [PTFP]) ermöglichen in die LTFU-001-Studie zu wechseln, wenn sie es wünschen.</li> <li>• Berücksichtigung der Coronavirus-Erkrankung (COVID-19) Erhebung von COVID-19-Daten. Ziel dieser Änderung ist es die Auswirkungen von COVID-19 auf den Studienverlauf zu bewerten (gemäß BMS-Richtlinien zum Umgang mit COVID-19-Daten im Rahmen von klinischen Studien).</li> <li>• Extramedulläre Blutbildung Ziel dieser Änderung ist die enge Beobachtung extramedullärer Blutbildung als unerwünschtes Ereignis von besonderem Interesse für die Sicherheit der Probanden</li> </ul>
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p><b>Einschlusskriterien:</b></p> <ul style="list-style-type: none"> <li>• Alter <math>\geq 18</math> Jahre bei Einverständniserklärung</li> <li>• Verständnis und freiwillige Gabe der Einverständniserklärung vor jeglichen studienbezogenen Untersuchungen/Prozeduren</li> <li>• Fähigkeit und Willen zur Einhaltung der Studienzeitpläne und anderen Anforderungen des Protokolls (z. B. keine hämatopoetische Stammzelltransplantation geplant)</li> <li>• dokumentierte Diagnose einer Beta-Thalassämie oder HbE/Beta-Thalassämie (Beta-Thalassämie mit Mutation und/oder Multiplikation des Alpha-Globins war erlaubt)</li> <li>• Nicht-Transfusionsabhängigkeit (NTD), definiert durch <ul style="list-style-type: none"> <li>○ Transfusionsbedarf von 0 – 5 Erythrozytenkonzentrat (EK)-Einheiten in den 24</li> </ul> </li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Wochen vor Randomisierung (eine Einheit wurde definiert als 200 bis 350 ml EK) und</p> <ul style="list-style-type: none"> <li>○ Transfusionsfreie Periode von mindestens <math>\geq 8</math> Wochen vor Randomisierung</li> <li>• Keine Teilnahme an einem regulären Transfusionsprogramm</li> <li>• Mittlerer Hämoglobin-Wert zu Baseline von <math>\leq 10,0</math> g/dl, basierend auf mindestens 2 Messungen, mit <math>\geq 1</math> Woche Abstand innerhalb von 4 Wochen vor Randomisierung (Hämoglobin-Werte innerhalb von 21 Tagen nach Transfusion wurden ausgeschlossen)</li> <li>• ECOG-Status von 0 oder 1</li> <li>• Frauen im gebärfähigen Alter mussten <ul style="list-style-type: none"> <li>○ 2 negative, durch den Prüfarzt bestätigte Schwangerschaftstests vor Beginn der Studienmedikation vorliegen haben und ihr Einverständnis geben, während der Studie und nach Behandlungsende weitere Schwangerschaftstest durchzuführen (auch wenn die Patient:innen von heterosexuellen Kontakten abstinent waren)</li> <li>○ entweder vollständige Abstinenz von heterosexuellen Kontakten wahren oder, im Fall von sexuellen Aktivitäten, die zu einer Schwangerschaft führen könnten, zustimmen und in der Lage sein, eine hoch wirksame Methode der Kontrazeption ohne Unterbrechung ab 28 Tage vor Behandlungsbeginn, während der Behandlung mit Studienmedikation und für 12 Wochen nach Beendigung der Studienmedikation einzuhalten (definiert als Methode mit einem Pearl Index <math>&lt; 1</math> % pro Jahr: (1) orale, intravaginale und transdermale kombinierte (Estrogen und Progesteron) hormonelle Kontrazeption, (2) orale, injizierbare oder implantierbare alleinige hormonelle Kontrazeption mit Progesteron und damit verbundener Ovulationsinhibierung, (3) Intrauterin-pessar, (4) intrauterines Hormon-abgebendes System, (5) bilaterale Tubenligatur, (6) ein vasktomierter Partner und (7) sexuelle Abstinenz)</li> </ul> </li> <li>• Männliche Probanden mussten entweder vollständige Abstinenz wahren oder einwilligen, während der Studienteilnahme, Dosisunterbrechungen und für mindestens 12 Wochen nach Abbruch der Studienmedikation bei Geschlechtsverkehr mit schwangeren Frauen oder Frauen im gebärfähigen Alter Kondome zu verwenden – selbst bei erfolgreich durchgeführter Vasektomie</li> </ul> <p><b>Ausschlusskriterien:</b></p> <ul style="list-style-type: none"> <li>• Jeglicher bedeutende medizinische Zustand, Laborwert-anomalie oder psychiatrische Erkrankung, die die Teilnahme an der Studie verhinderten</li> <li>• Jegliche Kondition, einschließlich des Vorliegens von Laborwertanomalien, die den Patienten oder die Patientin</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>einem nicht akzeptablen Risiko aussetzen würde, sollte er/sie in der Studie teilnehmen</p> <ul style="list-style-type: none"> <li>• Jegliche Kondition oder begleitende Medikation, die in der Lage war, die Interpretierbarkeit der Studiendaten zu beeinträchtigen</li> <li>• Diagnose einer HbS/Beta-Thalassämie oder Alpha-Thalassämie (z.B. HbH)</li> <li>• Aktive Hepatitis C-Virus (HCV)-Infektion, nachgewiesen durch positiven HCV-RNA-Test mit ausreichender Sensitivität, oder aktive infektiöse HBV-Infektion, nachgewiesen durch HBV-Oberflächenantigene (HBsAg) und/oder positive HBV-DNA, oder bekannte positive HIV-Infektion (Patient:innen, die antiviral behandelt wurden, sollten vor der Einverständniserklärung zwei negative HCV-RNA-Tests im Abstand von 3 Monaten Abstand vorweisen)</li> <li>• Tiefe Venenthrombose oder Schlaganfall mit erforderlicher medizinischer Intervention innerhalb der 24 Wochen vor Randomisierung</li> <li>• Chronische Behandlung mit Antikoagulantien, außer die Behandlung wurde mindestens 28 Tage vor Randomisierung beendet; erlaubt waren die Antikoagulantien zu Prophylaxe, vor Operationen oder Hochrisikoeingriffen, sowie niedrigmolekularem Heparin zur Behandlung von oberflächlichen Venenthrombosen und Aspirin zur Behandlung chronischer Erkrankungen</li> <li>• Behandlung mit anderer Prüfsubstanz oder -gerät <math>\leq 28</math> Tage vor Randomisierung</li> <li>• Vorherige Behandlung mit Luspatercept (ACE-536) oder Sotatercept (ACE-011)</li> <li>• Thrombozytenzahl <math>&gt;1000 \times 10^9/l</math></li> <li>• Proband:innen, die zur Zeit der Einverständniserklärung eine Eisenchelatherapie (ICT) erhielten, musste die ICT-Behandlung mindestens 24 Wochen vor dem erwarteten Datum der Randomisierung begonnen haben. Die ICT konnte zu jedem Zeitpunkt der Behandlung mit der Studienmedikation begonnen werden und war entsprechend des Labels anzuwenden.</li> <li>• Behandlung mit ESA und/oder Hydroxyurea <math>\leq 24</math> Wochen vor Randomisierung sowie vorhergehende Genterapie</li> <li>• Schwangere oder stillende Frauen</li> <li>• Unkontrollierte Hypertonie; kontrollierte Hypertonie war definiert als <math>\leq</math> Grad 1 gemäß NCI CTCAE Version 4.0</li> <li>• Bedeutende Organschädigung, darunter <ul style="list-style-type: none"> <li>○ Lebererkrankung mit Alaninaminotransferase (ALT) <math>&gt; 3 \times</math> dem oberen Normalwert (<i>upper limit of normal</i>, ULN) oder Vorgeschichte einer nachgewiesenen Zirrhose sowie das Vorhandensein von soliden Leber-Massen/Tumore, die während des Screenings per Ultraschall festgestellt wurden</li> <li>○ Herzerkrankung, Herzinsuffizienz gemäß New York Heart Association Klassifizierung <math>\geq</math> Grad 3 oder</li> </ul> </li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>bedeutende, behandlungsbedürftige Arrhythmie oder kürzlicher Herzinfarkt innerhalb der 6 Monate vor Randomisierung</p> <ul style="list-style-type: none"> <li>○ Schwere Lungenerkrankung, einschließlich pulmonaler Fibrose oder pulmonaler Hypertonie, d. h. NCI CTCAE Version 4.0 <math>\geq</math> Grad 3</li> <li>○ Geschätzte glomeruläre Filtrationsrate <math>&lt; 60</math> ml/min/1,73 m<sup>2</sup> (mittels <i>Modification of Diet in Renal Disease</i>-Formel berechnet)</li> </ul> <ul style="list-style-type: none"> <li>• Chronische Behandlung mit systemischen Glucocorticoiden innerhalb der 12 Wochen vor Randomisierung (physiologische Ersatztherapie für Nebenniereninsuffizienz sowie einzelne Tage mit Glucocorticoid-Gabe waren erlaubt)</li> <li>• Größere Operation innerhalb der 12 Wochen vor Randomisierung; Patient:innen müssen sich vollständig von jeglichen vorangegangenen Operationen vor Randomisierung erholt haben</li> <li>• Vorgeschichte einer schweren allergischen oder anaphylaktischen Reaktion oder Hypersensitivität gegenüber rekombinanten Proteinen oder Inhaltsstoffen der Prüfsubstanz</li> <li>• Verwendung von Immunsuppressiva innerhalb von 28 Tagen vor der Randomisierung</li> <li>• Vorhergehende oder bestehende bösartige Erkrankungen (solide Tumore und hämatologische Malignome), es sei denn, der Proband/die Probandin war seit <math>\geq 5</math> Jahren erkrankungsfrei (einschließlich Abschluss einer aktiven oder adjuvanten Behandlung für eine frühere bösartige Erkrankung). Folgende Erkrankungen waren kein Ausschlussgrund: Basal- oder Plattenepithelkarzinom der Haut, Carcinoma in situ des Gebärmutterhalses, Carcinoma in situ der Brust, histologischer Zufallsbefund von Prostatakrebs (T1a oder T1b unter Verwendung der Tumor-, Nodus-, Metastasen-Klassifikation)</li> </ul>
4b	Studienorganisation und Ort der Studiendurchführung	<p><b>Orte der Studiendurchführung:</b> 12 Zentren und 6 Länder (Griechenland, Italien, Libanon, Thailand, Vereinigtes Königreich, USA)</p>
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	<p><b>Luspatercept:</b> Initialdosis: 1 mg/kg Körpergewicht s.c. q3w (Verabreichung an Tag 1 jedes 21-Tage-Behandlungszyklus) Maximale Gesamtdosis/Gabe: 120 mg Maximaldosis: 1,25 mg/kg Körpergewicht Dosisreduktion: um ~25%/Verabreichung (bis 0,45 mg/kg Körpergewicht)</p> <p><b>Placebo:</b> Volumenäquivalent zu Luspatercept s.c. q3w</p> <p><b>in beiden Armen:</b></p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		+ BSC in der DBTP und der OLP (darunter EK-Transfusionen, Eisenchelattherapie, Anämiebehandlung, Antibiotika, Virostatika, Antimykotika, Nahrungsergänzung), sofern klinisch angezeigt
<b>6</b>	Zielkriterien	
<b>6a</b>	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><b>Primärer Endpunkt:</b></p> <ul style="list-style-type: none"> <li>• Anteil der Proband:innen mit einer Erhöhung der mittleren Hämoglobin-Konzentration im Vergleich zu Baseline um <math>\geq 1,0</math> g/dl in Abwesenheit von Transfusionen über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13 bis 24)</li> </ul> <p><b>Wesentliche sekundäre Endpunkte:</b></p> <ul style="list-style-type: none"> <li>• Mittlere Änderung der NTDT-PRO Gesamtscores in der Domäne Müdigkeit/Schwäche (T/W) über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13 bis 24) im Vergleich zur Baseline</li> <li>• Mittlere Änderung der mittleren Hämoglobin-Werte über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13-24) in Abwesenheit von Transfusionen, im Vergleich zur Baseline</li> <li>• Anteil der Proband:innen mit einer Erhöhung der mittleren Hämoglobin-Konzentration von <math>\geq 1,0</math> g/dl über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 37 bis 48) in Abwesenheit von Transfusionen, im Vergleich zur Baseline</li> </ul> <p><b>Weitere Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Mittlere Änderung des durchschnittlichen Wertes der FACIT-Fatigue Skala über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13-24) im Vergleich zur Baseline</li> <li>• Mittlere Änderung des durchschnittlichen Wertes des NTDT-PRO Gesamtscores in der Domäne Kurzatmigkeit (SoB) über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13-24) im Vergleich zur Baseline</li> <li>• Mittlere Änderung des durchschnittlichen Hämoglobin-Werts über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 37-48) in Abwesenheit von Transfusionen, im Vergleich zur Baseline</li> <li>• Mittlere Änderung der durchschnittlichen Werte der FACIT-Fatigue Skala und der NTDT-PRO Gesamtscores in den Domänen Müdigkeit/Schwäche (T/W) Kurzatmigkeit (SoB) über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 37-48) im Vergleich zur Baseline</li> <li>• Anteil der Proband:innen mit einer Erhöhung der durchschnittlichen FACIT-Fatigue Skala von <math>\geq 3</math> Punkten über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13-24) im Vergleich zur Baseline</li> <li>• Anteil der Proband:innen mit einer Erhöhung der durchschnittlichen FACIT-Fatigue Skala von <math>\geq 3</math> über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 24 bis 48) im Vergleich zur Baseline</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> <li>• Mittlere Änderung der SF-36 PCS und MCS zu Woche 24 und Woche 48, im Vergleich zur Baseline</li> <li>• Anteil der Proband:innen mit Verbesserung der Eisenüberladung zu Woche 24 und 48 <ul style="list-style-type: none"> <li>○ Proband:innen mit LIC (durch MRT) <math>\geq 3</math> mg/g d.w. zur Baseline: <math>\geq 20\%</math> Reduktion in LIC oder <math>\geq 33\%</math> Abnahme der täglichen Dosis Eisenchelator</li> <li>○ Proband:innen mit LIC (durch MRT) <math>&lt; 2</math> mg/g d.w. zur Baseline: keine Zunahme des LIC <math>&gt; 1</math> mg/g d.w. und keine begonnene Behandlung mit Eisenchelator oder keine Zunahme der täglichen Dosis Eisenchelator <math>\geq 33\%</math> (bei Eisenchelatherapie zu Baseline)</li> </ul> </li> <li>• Mittlere Änderung des Serum-Ferritin-Wertes zu Woche 24, Woche 49 und zum Zeitpunkt der letzten Erhebung</li> <li>• Mittlere Änderung des LIC zu Woche 24, Woche 48 und zum Zeitpunkt der letzten Erhebung</li> <li>• Anteil der Proband:innen, die über 24 Wochen lang transfusionsfrei blieben</li> <li>• Anteil der Proband:innen, die über 48 Wochen lang transfusionsfrei blieben</li> <li>• Dauer der Erhöhung der mittleren Hämoglobin-Konzentration um <math>\geq 1,0</math> g/dl im Vergleich zur Baseline</li> <li>• Mittlere Änderung der 6MWT Distanz zu Woche 24 und 48 im Vergleich zur Baseline</li> <li>• Anteil der Proband:innen mit einer Erhöhung der mittleren Hämoglobin-Konzentration von <math>\geq 1,5</math> g/dl über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13-24) in Abwesenheit von Transfusionen, im Vergleich zur Baseline</li> <li>• Anteil der Proband:innen mit einer Abnahme des durchschnittlichen NTDT-PRO Gesamtscores in der Domäne Müdigkeit/Schwäche (T/W) um mindestens ein definiertes Responsekriterium, von Woche 13-24 und Woche 37-48, im Vergleich zur Baseline</li> </ul> <p><b>Sicherheit und Pharmakokinetik:</b></p> <ul style="list-style-type: none"> <li>• Sicherheit und Toleranz, inklusive Immunogenität</li> <li>• Pharmakokinetische Parameter</li> </ul> <p><b>Explorative Endpunkte:</b></p> <ul style="list-style-type: none"> <li>• Messung von Komorbiditäten der Erkrankung <ul style="list-style-type: none"> <li>○ Veränderung des extramedullären hämatopoetischen Massevolumens (durch MRT)</li> <li>○ Veränderung des Milzvolumens vorzugsweise durch MRT</li> <li>○ Veränderung der Trikuspidalklappen-Regurgitationsgeschwindigkeit durch Echokardiographie oder MRT</li> <li>○ Veränderung der Knochendichte durch Dual-Röntgen-Absorptiometrie zu Woche 48</li> </ul> </li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> <li>○ Falls vorhanden, durchschnittliche Veränderung der Größe von Beinulcera</li> <li>• Veränderung des NTDT Severity Score zu Woche 24 und 48 im Vergleich zur Baseline</li> <li>• Veränderung der Parameter zum morbiditätsfreien Überleben zu Woche 24, 48 und 144 im Vergleich zur Baseline</li> <li>• Mittlere Änderung des Serum GDF11 und anderer relevanter Biomarker-Korrelationen zur Hämoglobinantwort</li> <li>• Durchschnittliche Veränderungen des fetalen Hämoglobins (HbF)</li> <li>• Durchschnittliche Veränderung der Nutzung von Gesundheitsressourcen (HRU)</li> </ul>
<b>6b</b>	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Nicht zutreffend.
<b>7</b>	Fallzahl	145
<b>7a</b>	Wie wurden die Fallzahlen bestimmt?	<p>Basierend auf einer angenommenen Ansprechrate des primären Endpunktes von mindestens 50,0% im Luspatercept-Arm und 10,0% im Placebo-Arm und basierend auf der 2:1 Randomisierung, wurde eine Fallzahl von 150 Proband:innen (100 im Luspatercept-Arm und 50 im Placebo-Arm) für eine Power von 99% berechnet, um Unterschiede zwischen den beiden Behandlungsgruppen mit einem zweiseitigen <math>\alpha</math> von 0,05 festzustellen, wobei von einer 10,0% Dropout-Rate ausgegangen wurde.</p> <p>Hinsichtlich des NTDT-PRO Gesamtscores in der Domäne Müdigkeit/Schwäche (T/W) wurden zu Woche 24 eine mittlere Änderung von 1,2 bzw. 0,5 Punkten (Luspatercept bzw. Placebo) mit einer Standardabweichung (SD) von 1,2 Punkten angenommen. Die statistische Power, um Unterschiede festzustellen, wurde auf 91% geschätzt.</p>
<b>7b</b>	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Es waren keine Interimanalysen geplant.
<b>8</b>	Randomisierung, Erzeugung der Behandlungsfolge	
<b>8a</b>	Methode zur Generierung der zufälligen Zuteilung	Zentrale Randomisierung mittels Interactive Response Technology (IRT) System
<b>8b</b>	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	<p>Blockrandomisierung stratifiziert nach:</p> <p>Hämoglobin-Wert zu Baseline</p> <ul style="list-style-type: none"> <li>• <math>\geq 8,5</math> g/dl</li> <li>• <math>&lt; 8,5</math> g/dl</li> </ul> <p>NTDT-PRO-Gesamtscore in der Domäne Müdigkeit/Schwäche (T/W) zu Baseline</p> <ul style="list-style-type: none"> <li>• <math>\geq 3</math> Punkte</li> <li>• <math>&lt; 3</math> Punkte</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Ausgewähltem Personal an jedem Studienzentrum wurden Passwort-geschützte, kodierte Identifizierungsnummern zugeordnet, die zur Einwahl in das IRT-System autorisierten, um Patient:innen einzuschließen. Die Randomisierung erfolgte anhand eines Algorithmus durch das IRT-System. Zu Behandlungsbeginn und bei jeder Dosisreduktion oder -titration musste das Personal am Zentrum die neue Dosierung im IRT-System aufnehmen, um die richtige Medikation zu erhalten.
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?	Die Randomisierung erfolgte durch das IRT-System, welches Patient:innen fortlaufend entsprechend der präspezifizierten Kriterien in ein geeignetes Stratum einsortierte.
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	a) ja b) ja c) ja Die Verblindung wurde für alle Beteiligten (mit Ausnahme bestimmter Personen, wie z. B. der Apothekerin oder dem Apotheker im Studienzentrum) aufrecht erhalten, bis alle Patient:innen die 48-wöchige doppelblinde Studienphase beendet oder abgebrochen hatten bzw. bis der Datenbankschluss erfolgt war und die Studie gemäß DMC Empfehlung entblindet wurde.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Für den Kontrollarm wurde eine 0,9%ige Natriumchlorid-Lösung zur Injektion in einer exakt gleich aussehenden Spritze wie die, die für rekonstituiertes Luspatercept verwendet wurde, vom ausgewählten Studienpersonal (z. B. der Apothekerin oder dem Apotheker) aufgezogen, um ein passendes Placebo herzustellen. Die Abgabe von Luspatercept und Placebo an den Prüfarzt oder die Prüfarztin und dessen/deren Angestellte erfolgte verblindet.
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p><b>Populationen für die Wirksamkeitsendpunkte:</b></p> <ul style="list-style-type: none"> <li>• <b>Intention-to-treat (ITT)-Population:</b> alle randomisierten Patient:innen, unabhängig davon, ob die Studienmedikation erhalten wurde</li> <li>• <b>Per-Protocol-Set (PPS)-Population:</b> alle randomisierten Patient:innen, die mindestens 1 Dosis der Studienmedikation erhalten haben und bei denen keine größeren Abweichungen vom Protokoll festgestellt wurden</li> </ul> <p><b>Populationen für die Sicherheitsendpunkte:</b></p> <ul style="list-style-type: none"> <li>• <b>Safety Population:</b> alle randomisierten Patient:innen, die mindestens 1 Dosis der Studienmedikation erhalten haben; Patient:innen werden den Behandlungsarmen zugeordnet, dessen Medikation sie tatsächlich erhalten haben</li> </ul> <p><b>Populationen für die Endpunkte der Lebensqualität:</b></p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> <li>• <b>HRQoL Population (<i>HRQoL-evaluable population</i>):</b> alle Patient:innen der ITT-Population, die Lebensqualitätsfragebogen zur Baseline (d. h. Screening Visite) und mindestens 1 Post-Baseline-Untersuchung vollständig ausgefüllt haben</li> </ul> <p><b>Populationen für die Pharmakokinetik-Endpunkte:</b></p> <ul style="list-style-type: none"> <li>• <b>Pharmakokinetik-Population:</b> alle Patient:innen, die mindestens 1 Dosis Luspatercept erhalten haben und messbare Luspatercept-Serumkonzentrationen aufwiesen</li> </ul> <p><b>Multiplizität:</b></p> <p>Zur Kontrolle des Fehlers 1. Art erfolgte für die drei wesentlichen sekundären Endpunkte durch eine Gatekeeping-Prozedur: Der Test für den ersten sekundären Endpunkt wurde nur durchgeführt, wenn der primäre Endpunkt erreicht worden war. Der Test für den zweiten sekundären Endpunkt wurde nur durchgeführt, wenn der erste sekundäre Endpunkt erreicht worden war. Der Test für den dritten sekundären Endpunkt wurde nur durchgeführt, wenn der zweite sekundäre Endpunkt erreicht worden war.</p> <p>Signifikanzniveau jeweils 0,05.</p> <p><b>Analyse des primären Endpunkts:</b></p> <p>Das erythroide Ansprechen als primärer Wirksamkeitsendpunkt war definiert als der Anteil der Proband:innen mit einer Zunahme des durchschnittlichen Hämoglobin-Werts von <math>\geq 1,0</math> g/dl im Vergleich zur Baseline über einen kontinuierlichen Zeitraum von 12 Wochen (Woche 13-24) in Abwesenheit von Transfusionen. Hämoglobin-Werte innerhalb von 21 Tage nach einer Transfusion wurden von der Analyse ausgeschlossen. Der Baseline-Hämoglobin-Wert wurde definiert als Durchschnitt von <math>\geq 2</math> Messungen im Abstand von mindestens einer Woche innerhalb der 4 Wochen vor der ersten Dosis.</p> <p>Der primäre Endpunkt wurde anhand der ITT-Population ausgewertet.</p> <p>Die Auswertung des Anteils der Proband:innen mit einem erythroiden Ansprechen erfolgte mittels Cochran-Mantel-Haenszel-(CMH-)Test mit einem zweiseitigem Signifikanzniveau von 0,05, stratifiziert nach den Stratifikationsfaktoren der Randomisierung. Der Vergleich von Luspatercept und Placebo erfolgte anhand des Odds Ratios (OR).</p> <p><b>Analyse der sekundären Hauptendpunkte:</b></p> <p>Die mittlere Änderung der NTDT-PRO Gesamtscores in der Domäne Müdigkeit/Schwäche (T/W) über einen Zeitraum von 12 Wochen (Woche 13-24) im Vergleich zur Baseline wurde mittels ANCOVA berechnet.</p> <p>Die mittlere Änderung des Hämoglobin-Werts über einen Zeitraum von 12 Wochen (Woche 13-24) in Abwesenheit von Transfusionen im Vergleich zur Baseline wurde mittels ANCOVA berechnet.</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		Der Anteil der Proband:innen mit einer Zunahme des mittleren Hämoglobin-Werts von $\geq 1,0$ g/dl innerhalb eines Zeitraums von 12 Wochen (Woche 37-48) in Abwesenheit von Transfusionen im Vergleich zur Baseline wurde entsprechend analog zum primären Endpunkt ausgewertet.
<b>12b</b>	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Durchführung von Subgruppenanalysen für den primären und die wesentlichen sekundären Endpunkte nach folgenden Merkmalen: <ul style="list-style-type: none"> <li>• Hb</li> <li>• NTDT-PRO Gesamtscore in der Domäne Müdigkeit/Schwäche (T/W)</li> <li>• Geografische Region</li> <li>• Alter</li> <li>• Splenektomie</li> <li>• Geschlecht</li> <li>• Diagnose</li> <li>• LIC</li> <li>• 6MWT</li> <li>• FACIT-Fatigue Skala</li> </ul>
<b>Resultate</b>		
<b>13</b>	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
<b>13a</b>	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die <ul style="list-style-type: none"> <li>a) randomisiert wurden,</li> <li>b) tatsächlich die geplante Behandlung/Intervention erhalten haben,</li> <li>c) in der Analyse des primären Zielkriteriums berücksichtigt wurden</li> </ul>	Luspatercept + BSC vs. Placebo + BSC: <ul style="list-style-type: none"> <li>a) 96 vs. 49 (gesamt: 145)</li> <li>b) 96 vs. 49 (gesamt: 145)</li> <li>c) 96 vs. 49 (gesamt: 145)</li> </ul>
<b>13b</b>	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Siehe Abbildung 4-26
<b>14</b>	Aufnahme / Rekrutierung	
<b>14a</b>	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	First subject first visit: 05.03.2018 Last subject last visit: 28.11.2022
<b>14b</b>	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde wie geplant nach Abschluss der Open-Label-Phase und Wechsel aller verbliebenen Patient:innen in die

Item <sup>a</sup>	Charakteristikum	Studieninformation
		Rollover-Studie ACE-536-LTFU-001 beendet. Der finale CSR lag zum Zeitpunkt der Dossiereinreichung noch nicht vor.
<b>a: nach CONSORT 2010.</b>		

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

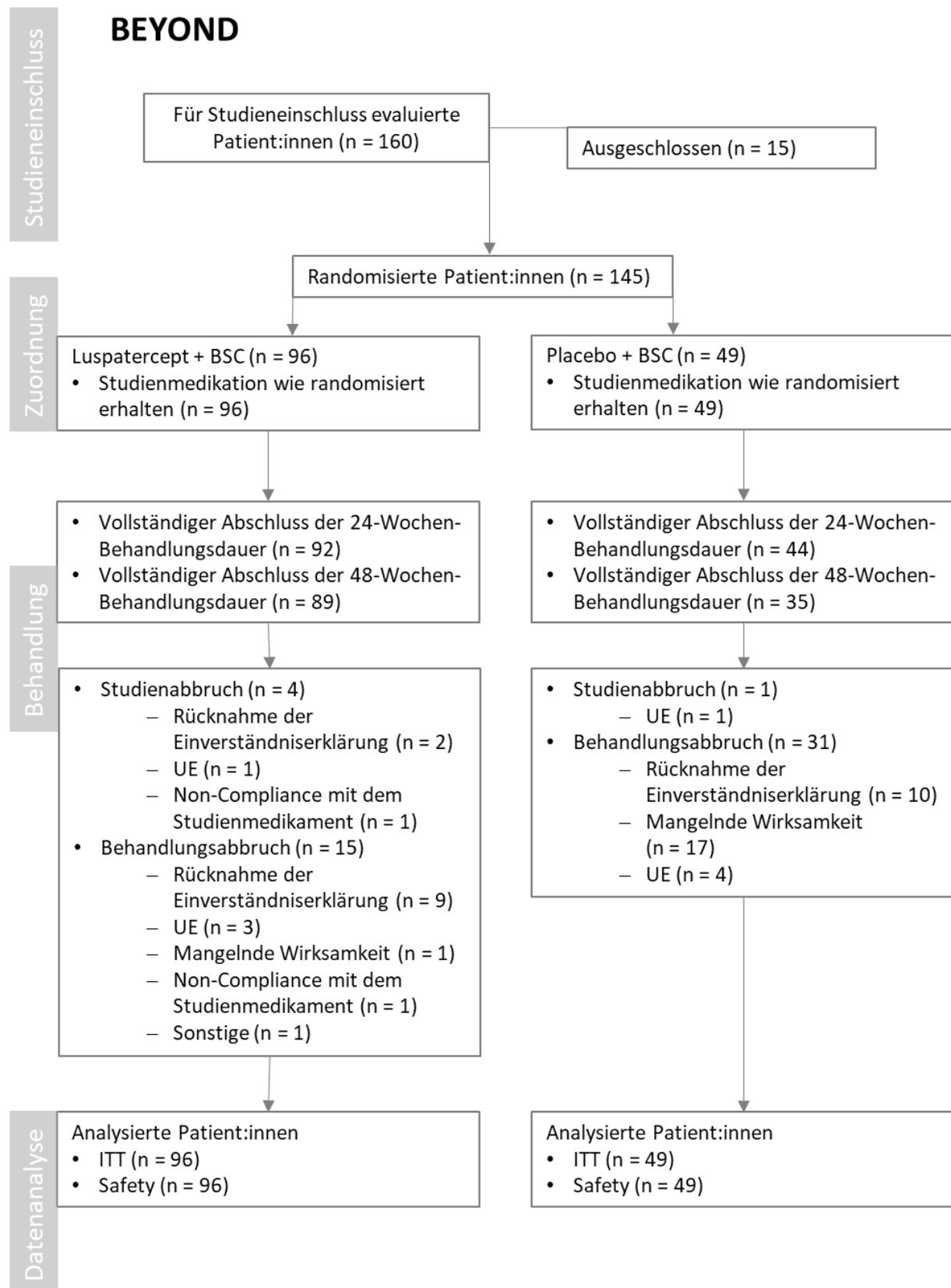


Abbildung 4-26: Patientenfluss in BEYOND bis zum Datenschnitt vom 14.09.2020

**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 4-86 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie BEYOND

**Studie: BEYOND**

**Tabelle: Liste der für die Bewertung herangezogenen Quellen**

Genauere Benennung der Quelle	Kürzel
Studienbericht	(A)
Studienprotokoll	(B)
Studienregistereintrag (Clinicaltrials.gov)	(C)
Studienregistereintrag (ICTRP)	(D)

**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:

Es handelt sich um eine randomisierte Studie.

A, B, C, D

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:

---

In dieser Studie waren alle Patient:innen, Studienpersonal und Vertreter:innen der Firma Celgene bzw. Bristol Myers Squibb hinsichtlich der Behandlung verblindet. Davon ausgenommen waren ausgewählte Einzelpersonen (z. B. Apotheker:innen des Studienzentrums), die eine adäquate Herstellung der Placebo-Gaben sicherstellten. Die Randomisierung, Arzneimittelausgabe, Dosisreduktion/-titrierung und Therapieabbrüche wurden durch das IRT-System unterstützt, sodass die Verblindung sichergestellt wurde. Die Behandlungszuweisung durfte nur entblindet werden, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient:innen zu gewähren. Eine Notfall-Entblindung musste anhand eines strikten Protokolls über das IRT-System erfolgen.

A, B

---

**behandelnde bzw. weiterbehandelnde Personen:**

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

In dieser Studie waren alle Patient:innen, Studienpersonal und Vertreter:innen der Firma Celgene bzw. Bristol Myers Squibb hinsichtlich der Behandlung verblindet. Davon ausgenommen waren ausgewählte Einzelpersonen (z. B. Apotheker:innen des Studienzentrums), die eine adäquate Herstellung der Placebo-Gaben sicherstellten. Die Randomisierung, Arzneimittelausgabe, Dosisreduktion/-titrierung und Therapieabbrüche wurden durch das IRT-System unterstützt, sodass die Verblindung sichergestellt wurde. Die Behandlungszuweisung durfte nur entblindet werden, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient:innen zu gewähren. Eine Notfall-Entblindung musste anhand eines strikten Protokolls über das IRT-System erfolgen.

A, B

---

**4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte**

ja       unklar       nein

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

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

---

**5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können** ja       neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Es gab keinerlei Hinweise auf sonstige Aspekte, die zu einer Verzerrung führen könnten.  
A, B, C, D

---

**Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):** niedrig       hoch

Begründung für die Einstufung:

---

Die Randomisierung erfolgte zentral unter Verwendung eines IRT-Systems. Während der gesamten Studie waren Studienteilnehmer:innen, -personal und -sponsor hinsichtlich der Behandlungszuteilung verblindet. Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige Verzerrung konnten nicht identifiziert werden.

Das Verzerrungspotenzial auf Studienebene wird somit als niedrig eingestuft.

---

**B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:****Endpunkt: Mortalität****1. Verblindung der Endpunkterheber**

ja     unklar     nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.

A, B

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja     unklar     nein

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

---

In die Analyse des Endpunktes gingen alle randomisierten Patient:innen entsprechend der ITT-Population ein.

A, B

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja     unklar     nein

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

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja     nein

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

---

Es gab keinerlei Hinweise auf sonstige Aspekte, die zu einer Verzerrung führen könnten.

A, B, C, D

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig     hoch

Begründung für die Einstufung:

---

Der Endpunkt „Mortalität“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung oder sonstige Aspekte, die zu einer Verzerrung führen könnten, liegen nicht vor. Der Eintritt des Todes jeglicher Ursache ist ein objektives Kriterium, dessen Feststellung frei von Subjektivität erfolgt.  
Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

---

## Endpunkt: Morbidität – Vermeidung von Transfusionen

### 1. Verblindung der Endpunkterheber

ja     unklar     nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient:innen zu gewähren.

A, B

---

### 2. Adäquate Umsetzung des ITT-Prinzips

ja     unklar     nein

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

---

In die Analyse des Endpunktes gingen alle randomisierten Patient:innen entsprechend der ITT-Population ein.

A, B

---

### 3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja     unklar     nein

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

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

---

### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja     nein

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

Es gab keine sonstigen relevanten Aspekte, die zu einer Verzerrung führen können.  
A, B

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

**niedrig**                       **hoch**

Begründung für die Einstufung:

---

Der Endpunkt „Vermeidung von Transfusionen“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch eine Ereigniszeitanalyse adressiert. Die Auswertung umfasst den Behandlungszeitraum bis einschließlich 20 Tage nach der letzten Dosis. In Bezug auf Patient:innen, die die Therapie abbrachen, ist bei der betrachteten Ereigniszeitanalyse grundsätzlich die Möglichkeit potenziell informativer Zensurierungen nicht auszuschließen. In Anbetracht der Größe des beobachtenden Effekts der Ereigniszeitanalyse sowie der konsistenten Responder-analysen, ist jedoch davon auszugehen, dass dies im Ergebnis keine entscheidende Rolle spielt.

Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

---

**Endpunkt: Morbidität – Kontrolle der Eisenüberladung**

**1. Verblindung der Endpunkterheber**

**ja**                       **unklar**                       **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studien-zentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.

A, B

---

**2. Adäquate Umsetzung des ITT-Prinzips**

**ja**                       **unklar**                       **nein**

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

---

In die Analyse des Endpunktes gingen alle randomisierten Patienten entsprechend der ITT-Population ein (sofern zu den erforderlichen Zeitpunkten Messungen des sFerritin vorlagen).

A, B

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

**ja**                       **unklar**                       **nein**

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

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.  
A, B, C, D

---

#### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja       nein

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

Es gab keinerlei Hinweise auf sonstige Aspekte, die zu einer Verzerrung führen könnten.  
A, B, C, D

---

#### Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig       hoch

Begründung für die Einstufung:

Der Endpunkt „Kontrolle der Eisenüberladung“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung oder sonstige Aspekte, die zu einer Verzerrung führen könnten, liegen nicht vor.  
Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

---

#### Endpunkt: Morbidität – Symptomatik gemäß NTDT-PRO

##### 1. Verblindung der Endpunkterheber

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüffärztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.  
A, B

---

##### 2. Adäquate Umsetzung des ITT-Prinzips

ja       unklar       nein

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

In die Analyse des Endpunktes gingen alle randomisierten Patienten entsprechend der ITT-Population ein (sofern zu den erforderlichen Zeitpunkten ausgefüllte Fragebögen vorlagen). Während zu Woche 24 hinreichend hohe Rücklaufquoten vorlagen (80,2 % im Luspatercept-Arm bzw. 79,6 % Placebo-Arm), sanken die Rücklaufquoten zu Woche 48 zwar unter die Marke von 70 % bezogen auf die ITT-Population (50,0 % bzw. 55,1 %), die Effekte der Responderanalysen zu Woche 24 und 48 blieben jedoch konsistent.  
A, B

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja     unklar     neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.  
A, B, C, D

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja     neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen.  
A, B

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig     hoch

Begründung für die Einstufung:

---

Der Endpunkt „Symptomatik gemäß NTDT-PRO“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert. Bei Patient:innen, die die Therapie abbrachen, wurde der NTDT-PRO weiterhin bis mindestens Woche 48 sowie zu einem Zeitpunkt 9 Wochen nach der letzten Dosis erhoben. Während zu Woche 24 hinreichend hohe Rücklaufquoten vorlagen (80,2 % im Luspatercept-Arm bzw. 79,6 % Placebo-Arm), sanken die Rücklaufquoten zu Woche 48 zwar unter die Marke von 70 % bezogen auf die ITT-Population (50,0 % bzw. 55,1 %), die Effekte der Responderanalysen zu Woche 24 und 48 blieben jedoch konsistent. In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

---

**Endpunkt: Morbidität – Symptomatik gemäß PGI-S und PGI-C****1. Verblindung der Endpunkterheber** ja     unklar     neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen,

---

die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.

A, B

---

## 2. Adäquate Umsetzung des ITT-Prinzips

ja       unklar       nein

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

---

In die Analyse des Endpunktes gingen alle randomisierten Patienten entsprechend der ITT-Population ein (sofern zu den erforderlichen Zeitpunkten ausgefüllte Fragebögen vorlagen). Für den PGI-C ergaben sich sowohl zu Woche 24, als auch zu Woche 48 hohe Rücklaufquoten (über 70 % in beiden Armen). Für den PGI-S waren die Rücklaufquoten zu Woche 24 ebenfalls hinreichend hoch (80,2 % im Luspatercept-Arm bzw. 79,6 % im Placebo-Arm); zwar sanken die Rücklaufquoten für den PGI-S zu Woche 48 unter die Marke von 70 % bezogen auf die ITT-Population (50,0 % bzw. 55,1 %), jedoch waren die Ergebnisse der Responderanalysen zu Woche 24 und 48 konsistent.

A, B

---

## 3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja       unklar       nein

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

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

---

## 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja       nein

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

---

Es gab keine sonstigen relevanten Aspekte, die zu einer Verzerrung führen können.

A, B

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig       hoch

Begründung für die Einstufung:

Der Endpunkt „Symptomatik gemäß PGI-S und PGI-C“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert.

Bei Patient:innen, die die Therapie abbrachen, wurden der PGI-S und PGI-C weiterhin bis mindestens Woche 48 sowie 9 Wochen nach der letzten Dosis erhoben.

Für den PGI-C ergaben sich sowohl zu Woche 24, als auch zu Woche 48 hohe Rücklaufquoten (über 70 % in beiden Armen). Für den PGI-S waren die Rücklaufquoten zu Woche 24 ebenfalls hinreichend hoch (80,2 % im Luspatercept-Arm bzw. 79,6 % im Placebo-Arm); zwar sanken die Rücklaufquoten für den PGI-S zu Woche 48 unter die Marke von 70 % bezogen auf die ITT-Population (50,0 % bzw. 55,1 %), jedoch waren die Ergebnisse der Responderanalysen zu Woche 24 und 48 konsistent.

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder die Fragebögen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensurierungen nicht auszuschließen. In Anbetracht der Größe der beobachtenden Effekte der Ereigniszeitanalysen sowie der konsistenten Responderanalysen, ist jedoch davon auszugehen, dass dies im Ergebnis keine entscheidende Rolle spielt.

Insgesamt wird das Verzerrungspotenzial beim PGI-S und PGI-C vor diesem Hintergrund als niedrig eingestuft.

### Endpunkt: Morbidität – körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest (6MWT)

#### 1. Verblindung der Endpunkterheber

ja     unklar     nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüffärztin absolut notwendig war, um eine sichere Behandlung der Patient:innen zu gewähren.

A, B

#### 2. Adäquate Umsetzung des ITT-Prinzips

ja     unklar     nein

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

In die Analyse des Endpunktes gingen alle randomisierten Patienten entsprechend der ITT-Population ein (sofern zu den erforderlichen Zeitpunkten Messungen des 6MWT vorlagen).

A, B

#### 3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja     unklar     nein

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

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja       neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Es gab keinerlei Hinweise auf sonstige Aspekte, die zu einer Verzerrung führen könnten.  
A, B, C, D

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig       hoch

Begründung für die Einstufung:

---

Der Endpunkt „Körperliche Leistungsfähigkeit gemäß 6-Minuten-Gehtest“ wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung oder sonstige Aspekte, die zu einer Verzerrung führen könnten, liegen nicht vor.  
Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt somit als niedrig eingestuft.

---

**Endpunkt: Lebensqualität – SF-36****1. Verblindung der Endpunkterheber** ja       unklar       neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.  
A, B

---

**2. Adäquate Umsetzung des ITT-Prinzips** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

In die Analyse des Endpunktes gingen alle randomisierten Patienten entsprechend der ITT-Population ein (sofern zu den erforderlichen Zeitpunkten ausgefüllte Fragebögen vorlagen). Für den SF-36 liegen sowohl für Woche 24, als auch für Woche 48 hohe Rücklaufquoten vor (über 70 % in beiden Behandlungsarmen).  
A, B

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja       unklar       nein

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

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.  
A, B, C, D

---

#### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja       nein

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

---

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen.  
A, B

---

#### Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig       hoch

Begründung für die Einstufung:

---

Der Endpunkt SF-36 wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert.  
Bei Patient:innen, die die Therapie abbrachen, wurde der SF-36 nur bis Woche 48 bzw. 9 Wochen nach der letzten Dosis erhoben.  
Für den SF-36 liegen sowohl für Woche 24, als auch für Woche 48 hohe Rücklaufquoten vor (über 70 % in beiden Behandlungsarmen).  
In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

---

#### Endpunkt: Lebensqualität – FACIT-F

##### 1. Verblindung der Endpunkterheber

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüffärztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.  
A, B

---

**2. Adäquate Umsetzung des ITT-Prinzips** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

In die Analyse des Endpunktes gingen alle randomisierten Patienten entsprechend der ITT-Population ein (sofern zu den erforderlichen Zeitpunkten ausgefüllte Fragebögen vorlagen). Für den FACIT-F liegen sowohl für Woche 24, als auch für Woche 48 hohe Rücklaufquoten vor (über 70 % in beiden Behandlungsarmen).

A, B

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja       neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen.

A, B, C, D

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig       hoch

Begründung für die Einstufung:

---

Der Endpunkt FACIT-F wurde verblindet erhoben. Das ITT-Prinzip wurde adäquat umgesetzt. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor.

Unterschiede in den Beobachtungszeiten wurden durch Ereigniszeitanalysen adressiert.

Bei Patient:innen, die die Therapie abbrachen, wurde der FACIT-F nur bis Woche 48 bzw. 9 Wochen nach der letzten Dosis erhoben.

Für den FACIT-F liegen sowohl für Woche 24, als auch für Woche 48 hohe Rücklaufquoten vor (über 70 % in beiden Behandlungsarmen).

In Bezug auf Patient:innen, die die Therapie abbrachen und/oder den Fragebogen nicht mehr ausfüllten, ist bei den betrachteten Ereigniszeitanalysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungs-potenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

---

**Endpunkt: Sicherheit****1. Verblindung der Endpunkterheber**

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

In der Studie BEYOND waren Patient:innen, Studienpersonal und Mitarbeiter:innen von Celgene bzw. Bristol Myers Squibb verblindet. Davon ausgenommen waren lediglich ausgewählte Einzelpersonen, die die korrekte Umsetzung der Placebogabe sicherstellten (z. B. Apotheker:innen des Studienzentrums). Eine Entblindung war nur gestattet, wenn dies aus Sicht des Prüfarztes / der Prüfarztin absolut notwendig war, um eine sichere Behandlung der Patient innen zu gewähren.

A, B

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja       unklar       nein

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

---

In die Analyse der UE gingen nur solche Patient:innen ein, die mindestens eine Dosis Studienmedikation erhalten haben (Safety-Population). In der vorliegenden Studie ergaben sich jedoch zwischen der Safety-Population und der ITT-Population keine Unterschiede.

A, B

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja       unklar       nein

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

---

Es gab keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung.

A, B, C, D

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja       nein

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

---

In Bezug auf Patient:innen, die die Therapie abbrechen, ist bei den betrachteten Ereigniszeit-analysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen.

A, B

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig       hoch

Begründung für die Einstufung:

Der Endpunkt wurde verblindet erhoben. In die Analyse der UE gingen nur solche Patient:innen ein, die mindestens eine Dosis Studienmedikation erhalten haben (Safety-Population). In der vorliegenden Studie ergaben sich jedoch zwischen der Safety-Population und der ITT-Population keine Unterschiede. Anhaltspunkte für eine ergebnisabhängige Berichterstattung liegen nicht vor. Entsprechend der Nachbeobachtung für unerwünschte Ereignisse von 63 Tagen nach der letzten Dosis, ergaben sich aus dem niedrigeren Anteil der Therapieabbrucher:innen und der längeren Therapiedauer unter Luspatercept im Vergleich zu Placebo unterschiedliche Beobachtungsdauern. Diese Unterschiede wurden durch Ereigniszeitanalysen adressiert. In Bezug auf Patient:innen, die die Therapie abbrachen, ist bei den betrachteten Ereigniszeit-analysen grundsätzlich die Möglichkeit potenziell informativer Zensierungen nicht auszuschließen. Insgesamt wird das Verzerrungspotenzial beim vorliegenden Endpunkt deshalb im Sinne einer konservativen Bewertung als hoch eingestuft.

---

Hinweis: Der nachfolgend dargestellte Bewertungsbogen mit Ausfüllhinweisen dient nur als Ausfüllhilfe für den Blankobogen. Er soll nicht als Vorlage verwendet werden.

### Bewertungsbogen zur Beschreibung von Verzerrungsaspekten (Ausfüllhilfe)

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunkübergreifend; B: endpunktspezifisch).

#### A Verzerrungsaspekte auf Studienebene:

##### Einstufung als randomisierte Studie

**ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien

**nein:** Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anzeichen vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

---

---

1.

##### für randomisierte Studien:

##### Adäquate Erzeugung der Randomisierungssequenz

**ja:** Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

**unklar:** Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

**nein:** Die Erzeugung der Zuteilungssequenz war nicht adäquat.

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

---

---

##### für nicht randomisierte Studien:

##### Zeitliche Parallelität der Gruppen

**ja:** Die Gruppen wurden zeitlich parallel verfolgt.

**unklar:** Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

**nein:** Die Gruppen wurden nicht zeitlich parallel verfolgt.

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

---

2.

**für randomisierte Studien:****Verdeckung der Gruppenzuteilung („allocation concealment“)** **ja:** Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

 **unklar:** Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau. **nein:** Die Gruppenzuteilung erfolgte nicht verdeckt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

**für nicht randomisierte Studien:****Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** **ja:** Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anzeichen dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

 **unklar:** Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau. **nein:** Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

**3. Verblindung von Patienten und behandelnden Personen****Patient:** **ja:** Die Patienten waren verblindet. **unklar:** Es finden sich keine diesbezüglichen Angaben. **nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---



---

**behandelnde bzw. weiterbehandelnde Personen:**

**ja:** Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, die primär behandelnde Person (z. B. Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

**unklar:** Es finden sich keine diesbezüglichen Angaben.

**nein:** Aus den Angaben geht hervor, dass die behandelnden Personen nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---



---

**4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte**

*Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).*

*Beispiele zu a und b:*

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

*Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter*

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off- 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 Analysen**

Auf den nachfolgenden Seiten finden sich die Zusatzanalysen für alle Endpunkte als Ausgabe der Statistik-Software.

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

Protocol: ACE-536-B-THAL-002

Page 1 of 6

Time to First Transfusion Subgroups Analysis  
ITT Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	14 ( 14.6)	N.A.	49	21 ( 42.9)	86.14 (43.86, N.A.)	0.22 ( 0.11, 0.45) <0.0001	
REGION								0.5833
NORTH AMERICA & EUROPE	60	5 ( 8.3)	N.A.	30	9 ( 30.0)	N.A. (43.86, N.A.)	0.18 ( 0.06, 0.61) 0.0019	
MIDDLE EAST	9	3 ( 33.3)	N.A. ( 9.86, N.A.)	8	5 ( 62.5)	84.86 (21.71, N.A.)	0.41 ( 0.10, 1.74) 0.2117	
ASIA-PACIFIC	27	6 ( 22.2)	N.A.	11	7 ( 63.6)	32.00 ( 3.43, N.A.)	0.25 ( 0.08, 0.75) 0.0074	
AGE								0.5650
<=32	35	5 ( 14.3)	N.A.	12	7 ( 58.3)	84.86 ( 3.71, N.A.)	0.17 ( 0.05, 0.55) 0.0009	
>32	61	9 ( 14.8)	N.A.	37	14 ( 37.8)	90.29 (43.86, N.A.)	0.25 ( 0.11, 0.60) 0.0008	

Time to first transfusion is defined as the time from the first dose to the first transfusion.  
 HR = hazard ratio; (1) KME = Kaplan-Meier estimate; N.M.E. = Not meaningful estimate; N.E. = Not estimable.  
 (2) Unstratified Cox proportional hazard model with treatment as only covariate. Hazard Ratio is Luspatercept over Placebo.  
 (3) Unstratified Log-rank Test to compare Luspatercept and Placebo. (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup.  
 Data after the efficacy cutoff date of a subject will not be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-ftransub-ubr700.sas 30MAY2022:07:58:47

Protocol: ACE-536-B-THAL-002

Page 2 of 6

Time to First Transfusion Subgroups Analysis  
ITT Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SPLENECTOMY								
YES	34	8 ( 23.5)	N.A.	26	15 ( 57.7)	54.00 (21.71, N.A.)	0.26 ( 0.11, 0.63) 0.0014	0.9856
NO	62	6 ( 9.7)	N.A.	23	6 ( 26.1)	N.A. (86.14, N.A.)	0.26 ( 0.08, 0.85) 0.0167	
SEX								
MALE	40	5 ( 12.5)	N.A.	23	9 ( 39.1)	N.A. (16.14, N.A.)	0.18 ( 0.06, 0.60) 0.0017	0.5638
FEMALE	56	9 ( 16.1)	N.A.	26	12 ( 46.2)	86.14 (31.86, N.A.)	0.26 ( 0.11, 0.63) 0.0013	

Time to first transfusion is defined as the time from the first dose to the first transfusion.  
 HR = hazard ratio; (1) KME = Kaplan-Meier estimate; N.M.E. = Not meaningful estimate; N.E. = Not estimable.  
 (2) Unstratified Cox proportional hazard model with treatment as only covariate. Hazard Ratio is Luspatercept over Placebo.  
 (3) Unstratified Log-rank Test to compare Luspatercept and Placebo. (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup.  
 Data after the efficacy cutoff date of a subject will not be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-ftransub-ibr700.sas 30MAY2022:07:58:47

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

Protocol: ACE-536-B-THAL-002

Page 3 of 6

Time to First Transfusion Subgroups Analysis  
ITT Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA DIAGNOSIS								0.7229
BETA-THALASSEMIA	63	7 ( 11.1)	N.A.	34	10 ( 29.4)	86.14 (54.43, N.A.)	0.21 ( 0.07, 0.59)	0.0013
HEMOGLOBIN E/BETA-THALASSEMIA	28	5 ( 17.9)	N.A.	11	7 ( 63.6)	32.00 ( 3.43, N.A.)	0.21 ( 0.06, 0.65)	0.0030
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	10 ( 14.5)	N.A.	33	16 ( 48.5)	84.86 (31.86, N.A.)	0.21 ( 0.10, 0.48)	0.6635
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	2 ( 33.3)	N.A. ( 0.29, N.A.)	4	4 (100.0)	18.93 ( 2.57, N.A.)	<0.0001 0.25 ( 0.04, 1.38)	0.0863
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	2 ( 9.5)	N.A.	12	1 ( 8.3)	N.A.	0.55 ( 0.03, 8.75)	0.6651

Time to first transfusion is defined as the time from the first dose to the first transfusion.  
 HR = hazard ratio; (1) KME = Kaplan-Meier estimate; N.M.E. = Not meaningful estimate; N.E. = Not estimable.  
 (2) Unstratified Cox proportional hazard model with treatment as only covariate. Hazard Ratio is Luspatercept over Placebo.  
 (3) Unstratified Log-rank Test to compare Luspatercept and Placebo. (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup.  
 Data after the efficacy cutoff date of a subject will not be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-ftransub-ebr700.sas 30MAY2022:07:58:47

Protocol: ACE-536-B-THAL-002

Page 4 of 6

Time to First Transfusion Subgroups Analysis  
ITT Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE HEMOGLOBIN <8.5 G/DL	55	11 ( 20.0)	N.A.	29	16 ( 55.2)	43.86 (21.14, N.A.)	0.24 ( 0.11, 0.52)	0.9456
>=8.5 G/DL	41	3 ( 7.3)	N.A.	20	5 ( 25.0)	N.A. (54.00, N.A.)	0.18 ( 0.04, 0.79)	0.0110
BASELINE NTDI-PRO T/W SCORE <3 POINTS	30	2 ( 6.7)	N.A.	14	6 ( 42.9)	86.14 (21.14, N.A.)	0.10 ( 0.02, 0.52)	0.2816
>=3 POINTS	66	12 ( 18.2)	N.A.	35	15 ( 42.9)	90.29 (32.00, N.A.)	0.28 ( 0.13, 0.62)	0.0009

Time to first transfusion is defined as the time from the first dose to the first transfusion.  
 HR = hazard ratio; (1) KME = Kaplan-Meier estimate; N.M.E. = Not meaningful estimate; N.E. = Not estimable.  
 (2) Unstratified Cox proportional hazard model with treatment as only covariate. Hazard Ratio is Luspatercept over Placebo.  
 (3) Unstratified Log-rank Test to compare Luspatercept and Placebo. (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup.  
 Data after the efficacy cutoff date of a subject will not be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-ftransub-eb700.sas 30MAY2022:07:58:47

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

Protocol: ACE-536-B-THAL-002

Page 5 of 6

Time to First Transfusion Subgroups Analysis  
ITT Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	3 ( 8.6)	N.A.	17	1 ( 5.9)	N.A.	1.49 ( 0.15, 14.31)	0.5613
>=3 TO <=5 MG/G DRY WEIGHT	23	1 ( 4.3)	N.A.	13	5 ( 38.5)	90.29 (23.43, N.A.)	0.7285 ( <0.01, 0.44)	
>5 TO <=7 MG/G DRY WEIGHT	12	3 ( 25.0)	N.A. (29.43, N.A.)	3	2 ( 66.7)	6.86 ( 2.57, N.A.)	0.0004 ( 0.04, 1.56)	
>7 TO <=15 MG/G DRY WEIGHT	14	4 ( 28.6)	N.A. (31.14, N.A.)	9	7 ( 77.8)	21.71 ( 3.43, 54.43)	0.1095 ( 0.06, 0.79)	
>15 MG/G DRY WEIGHT	11	3 ( 27.3)	N.A. (21.29, N.A.)	5	4 ( 80.0)	21.14 ( 1.29, N.A.)	0.23 ( 0.05, 1.00)	

Time to first transfusion is defined as the time from the first dose to the first transfusion.  
 HR = hazard ratio; (1) KME = Kaplan-Meier estimate; N.M.E. = Not meaningful estimate; N.E. = Not estimable.  
 (2) Unstratified Cox proportional hazard model with treatment as only covariate. Hazard Ratio is Luspatercept over Placebo.  
 (3) Unstratified Log-rank Test to compare Luspatercept and Placebo. (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup.  
 Data after the efficacy cutoff date of a subject will not be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-ftransub-ebr700.sas 30MAY2022:07:58:47

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

Protocol: ACE-536-B-THAL-002

Page 6 of 6

Time to First Transfusion Subgroups Analysis  
ITT Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	N	Subjects with Event n (%)	KME [95%CI] (wk) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT <=450 METERS	51	8 ( 15.7)	N.A.	31	14 ( 45.2)	86.14 (43.86, N.A.)	0.23 ( 0.10, 0.57)	0.7264
>450 METERS	45	6 ( 13.3)	N.A.	18	7 ( 38.9)	N.A. ( 6.86, N.A.)	0.24 ( 0.08, 0.74)	0.0073
BASELINE FACIT-F FS SCORE <43 POINTS	64	13 ( 20.3)	N.A.	32	14 ( 43.8)	84.86 (32.00, N.A.)	0.34 ( 0.16, 0.74)	0.0783
>=43 POINTS	32	1 ( 3.1)	N.A.	16	7 ( 43.8)	86.14 (23.43, N.A.)	0.04 (<0.01, 0.30)	<0.0001

Time to first transfusion is defined as the time from the first dose to the first transfusion.  
 HR = hazard ratio; (1) KME = Kaplan-Meier estimate; N.M.E. = Not meaningful estimate; N.E. = Not estimable.  
 (2) Unstratified Cox proportional hazard model with treatment as only covariate. Hazard Ratio is Luspatercept over Placebo.  
 (3) Unstratified Log-rank Test to compare Luspatercept and Placebo. (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup.  
 Data after the efficacy cutoff date of a subject will not be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-ftransub-eb700.sas 30MAY2022:07:58:47

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Page 1 of 2  
 Cutoff Date: 14SEP2020

Table 14.2.14.1  
 Mean Change in Derived Liver Iron Concentration (mg/g dry weight) from Baseline at Week 24/48  
 ITT Population

	Luspatercept (N=96)	Placebo (N=49)
Baseline [a]		
n	95	47
Mean	6.09	5.91
SD	6.151	5.823
Median	3.91	4.10
Q1, Q3	2.39, 7.35	1.80, 7.45
Min, Max	0.8, 39.9	0.7, 28.7
Week 24 [b]		
n	88	47
Mean	5.69	5.65
SD	5.959	5.699
Median	3.69	3.80
Q1, Q3	2.20, 6.69	1.80, 6.59
Min, Max	0.6, 40.5	1.0, 30.4
Mean Change from Baseline (Week 24) [b]		
n	87	45
Mean	-0.27	-0.18
SD	0.930	1.375
Median	0.00	0.00
Q1, Q3	-0.55, 0.20	-0.30, 0.49
Min, Max	-4.1, 1.4	-5.6, 2.2
LS Mean (SE) [b]	-0.40 (0.137)	-0.28 (0.175)
LS Mean of Difference (Luspatercept - Placebo) (95% CI) [b]	-0.12 (-0.52,0.28)	
P-value [b]	0.5547	

The value of LIC is either the value collected from eCRF or the value derived from T2\*, R2\* or R2 parameter depending on which techniques and software were used for MRI LIC acquisition. Subjects with LIC value >43 are not included in the analysis.  
 [a] Baseline is defined as the last value on or before the first dose of study drug is administered; Postbaseline is defined as the closest visit at Week 24 or Week 48.  
 [b] Estimates are based on an ANCOVA model with treatment group, baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$  g/dL), baseline NIDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ), and baseline LIC as covariates. The derived LIC value will be used for analysis. If a subject has any LIC value  $> 43$ , the subject's LIC value will be excluded from analysis.  
 Program path: /u02/projects/ace\_536/ace\_536\_bthal\_002/deliverables/d011\_csr/programs/tables/final/t\_lic\_sum\_chg.sas  
 Run date: 03MAY2021 Data source: ADSL ADBASE ADMRI Data Extraction date: 20NOV2020

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Page 2 of 2  
 Cutoff Date: 14SEP2020

Table 14.2.14.1  
 Mean Change in Derived Liver Iron Concentration (mg/g dry weight) from Baseline at Week 24/48  
 ITT Population

	Luspatercept (N=96)	Placebo (N=49)
Week 48 [b]		
n	82	42
Mean	5.98	5.34
SD	6.212	5.402
Median	4.04	3.97
Q1, Q3	2.16, 6.80	1.85, 6.76
Min, Max	0.9, 39.9	0.8, 29.4
Mean Change from Baseline (Week 48) [b]		
n	81	40
Mean	-0.26	-0.91
SD	1.540	2.950
Median	-0.06	-0.09
Q1, Q3	-0.63, 0.33	-1.01, 0.33
Min, Max	-6.0, 3.2	-14.9, 2.4
LS Mean (SE) [b]	-0.46 (0.265)	-1.13 (0.355)
LS Mean of Difference (Luspatercept - Placebo) (95% CI) [b]	0.66 (-0.15,1.47)	
P-value [b]	0.1071	

The value of LIC is either the value collected from eCRF or the value derived from T2\*, R2\* or R2 parameter depending on which techniques and software were used for MRI LIC acquisition. Subjects with LIC value >43 are not included in the analysis.  
 [a] Baseline is defined as the last value on or before the first dose of study drug is administered; Postbaseline is defined as the closest visit at Week 24 or Week 48.  
 [b] Estimates are based on an ANCOVA model with treatment group, baseline Hb category (>= 8.5 g/dL, < 8.5 g/dL), baseline NTDI-PRO T/W score category (>=3 vs. <3), and baseline LIC as covariates. The derived LIC value will be used for analysis. If a subject has any LIC value > 43, the subject's LIC value will be excluded from analysis.  
 Program path: /u02/projects/ace\_536/ace\_536\_bthal\_002/deliverables/d011\_csr/programs/tables/final/t\_lic\_sum\_chg.sas  
 Run date: 03MAY2021 Data source: ADSL ADBASE ADMRI Data Extraction date: 20NOV2020

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

Celgene Corporation  
Protocol: ACE-536-B-THAL-002

Page 1 of 2  
Cutoff Date: 14SEP2020

Table 14.2.11.2  
Mean Change in Mean Daily Dose (mg) of Iron Chelation Therapy by ICT Drug  
ITT Population

	Luspatercept (N=96)	Placebo (N=49)
ICT DRUG: DEFERASIROX		
Baseline [a]		
n	8	7
Mean	1403.82	739.23
SD	1807.264	449.545
Median	935.27	720.00
Q1, Q3	450.00, 1260.00	450.00, 1189.58
Min, Max	180.0, 5760.0	125.0, 1440.0
Post-baseline [b]		
n	8	7
Mean	742.54	493.44
SD	565.670	441.032
Median	543.21	360.00
Q1, Q3	360.00, 1220.00	201.43, 512.68
Min, Max	42.9, 1651.1	125.0, 1440.0
Mean Change from Baseline		
n	8	7
Mean	-661.27	-245.78
SD	1445.932	348.190
Median	0.00	0.00
Q1, Q3	-745.36, 64.73	-518.57, 0.00
Min, Max	-4108.9, 180.0	-829.6, 62.7
LS Mean (SE) [C]	-439.91 (165.124)	-489.99 (239.712)
LS Mean of Difference (Luspatercept - Placebo) (95% CI) [C]	50.09 (-456.32, 556.50)	
P-value [C]	0.8300	

Only subjects who did not change ICT drug from baseline to post-baseline and only one ICT drug has been used are included.  
 [a] The baseline mean daily dose is calculated using the ICT dosage during the 24 weeks on or prior to first study drug treatment.  
 [b] Post-baseline mean daily dose is calculated during the last 24 weeks on or prior to the end date of the first 48 week treatment if efficacy cutoff is on or after 48 weeks; or during the last 24 weeks on or prior to the efficacy cutoff if efficacy cutoff is before 48 weeks, but on or after 24 weeks; otherwise missing. The efficacy cutoff is defined as min(death date, study discontinuation date, last dose date+ 20, 14SEP2020).  
 [c] Estimates are based on an ANCOVA model with treatment group, baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$  g/dL), baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ), and baseline ICT mean daily dose as covariates.  
 Program path: /u02/projects/ace\_536/ace\_536\_bthal\_002/deliverables/d011\_csr/programs/tables/final/t\_ict\_sum\_chg.sas  
 Run date: 03MAY2021 Data source: adsl adict adbase Data Extraction date: 20NOV2020

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

Celgene Corporation  
Protocol: ACE-536-B-THAL-002

Page 2 of 2  
Cutoff Date: 14SEP2020

Table 14.2.11.2  
Mean Change in Mean Daily Dose (mg) of Iron Chelation Therapy by ICT Drug  
ITT Population

	Luspatercept (N=96)	Placebo (N=49)
ICT DRUG: DEFERIPRONE		
Baseline [a]		
n	8	2
Mean	6375.00	9625.00
SD	3777.282	3358.757
Median	4500.00	9625.00
Q1, Q3	4250.00, 9000.00	7250.00, 12000.00
Min, Max	2000.0, 13500.0	7250.0, 12000.0
Post-baseline [b]		
n	8	2
Mean	6230.65	10500.00
SD	2495.663	2121.320
Median	5172.62	10500.00
Q1, Q3	4250.00, 8250.00	9000.00, 12000.00
Min, Max	4000.0, 10500.0	9000.0, 12000.0
Mean Change from Baseline		
n	8	2
Mean	-144.35	875.00
SD	3609.090	1237.437
Median	0.00	875.00
Q1, Q3	-250.00, 750.00	0.00, 1750.00
Min, Max	-7654.8, 5500.0	0.0, 1750.0
LS Mean (SE) [C]	415.77 (1346.484)	5808.20 (3605.794)
LS Mean of Difference (Luspatercept - Placebo) (95% CI) [C]	-5392.43 (-13845.52, 3060.66)	
P-value [C]	0.1620	

Only subjects who did not change ICT drug from baseline to post-baseline and only one ICT drug has been used are included.  
 [a] The baseline mean daily dose is calculated using the ICT dosage during the 24 weeks on or prior to first study drug treatment.  
 [b] Post-baseline mean daily dose is calculated during the last 24 weeks on or prior to the end date of the first 48 week treatment if efficacy cutoff is on or after 48 weeks; or during the last 24 weeks on or prior to the efficacy cutoff if efficacy cutoff is before 48 weeks, but on or after 24 weeks; otherwise missing. The efficacy cutoff is defined as min(death date, study discontinuation date, last dose date+ 20, 14SEP2020).  
 [c] Estimates are based on an ANCOVA model with treatment group, baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$  g/dL), baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ), and baseline ICT mean daily dose as covariates.  
 Program path: /u02/projects/ace\_536/ace\_536\_bthal\_002/deliverables/d011\_csr/programs/tables/final/t\_ict\_sum\_chg.sas  
 Run date: 03MAY2021 Data source: adsl adict adbase Data Extraction date: 20NOV2020

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 5

Table 14.1.1.1  
 Summary of Completion Rate of NTDT-PRO  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
BASELINE	94/96 (97.9%)	49/49 (100.0%)	143/145 (98.6%)
WEEK 1	88/96 (91.7%)	45/49 (91.8%)	133/145 (91.7%)
WEEK 2	85/96 (88.5%)	45/49 (91.8%)	130/145 (89.7%)
WEEK 3	87/96 (90.6%)	48/49 (98.0%)	135/145 (93.1%)
WEEK 4	88/96 (91.7%)	45/49 (91.8%)	133/145 (91.7%)
WEEK 5	87/96 (90.6%)	46/49 (93.9%)	133/145 (91.7%)
WEEK 6	90/96 (93.8%)	45/49 (91.8%)	135/145 (93.1%)
WEEK 7	82/96 (85.4%)	43/49 (87.8%)	125/145 (86.2%)
WEEK 8	89/96 (92.7%)	46/49 (93.9%)	135/145 (93.1%)
WEEK 9	89/96 (92.7%)	44/49 (89.8%)	133/145 (91.7%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least half of the 6 weekly item scores for that week (visit) non-missing (i.e., >=3 weekly items).  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 5

Table 14.1.1.1  
 Summary of Completion Rate of NTDT-PRO  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 10	87/96 (90.6%)	44/49 (89.8%)	131/145 (90.3%)
WEEK 11	89/96 (92.7%)	40/49 (81.6%)	129/145 (89.0%)
WEEK 12	85/96 (88.5%)	41/49 (83.7%)	126/145 (86.9%)
WEEK 13	87/96 (90.6%)	40/49 (81.6%)	127/145 (87.6%)
WEEK 14	85/96 (88.5%)	42/49 (85.7%)	127/145 (87.6%)
WEEK 15	81/96 (84.4%)	41/49 (83.7%)	122/145 (84.1%)
WEEK 16	82/96 (85.4%)	40/49 (81.6%)	122/145 (84.1%)
WEEK 17	82/96 (85.4%)	42/49 (85.7%)	124/145 (85.5%)
WEEK 18	85/96 (88.5%)	43/49 (87.8%)	128/145 (88.3%)
WEEK 19	82/96 (85.4%)	43/49 (87.8%)	125/145 (86.2%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least half of the 6 weekly item scores for that week (visit) non-missing (i.e., >=3 weekly items).  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 5

Table 14.1.1.1  
 Summary of Completion Rate of NTD-T-PRO  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 20	83/96 (86.5%)	41/49 (83.7%)	124/145 (85.5%)
WEEK 21	87/96 (90.6%)	38/49 (77.6%)	125/145 (86.2%)
WEEK 22	82/96 (85.4%)	40/49 (81.6%)	122/145 (84.1%)
WEEK 23	80/96 (83.3%)	36/49 (73.5%)	116/145 (80.0%)
WEEK 24	77/96 (80.2%)	39/49 (79.6%)	116/145 (80.0%)
WEEK 30	59/96 (61.5%)	31/49 (63.3%)	90/145 (62.1%)
WEEK 36	58/96 (60.4%)	33/49 (67.3%)	91/145 (62.8%)
WEEK 42	53/96 (55.2%)	23/49 (46.9%)	76/145 (52.4%)
WEEK 48	48/96 (50.0%)	27/49 (55.1%)	75/145 (51.7%)
WEEK 54	38/96 (39.6%)	16/49 (32.7%)	54/145 (37.2%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least half of the 6 weekly item scores for that week (visit) non-missing (i.e., >=3 weekly items).  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 5

Table 14.1.1.1  
 Summary of Completion Rate of NTDT-PRO  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 60	39/96 (40.6%)	15/49 (30.6%)	54/145 (37.2%)
WEEK 66	39/96 (40.6%)	9/49 (18.4%)	48/145 (33.1%)
WEEK 72	41/96 (42.7%)	13/49 (26.5%)	54/145 (37.2%)
WEEK 78	35/96 (36.5%)	10/49 (20.4%)	45/145 (31.0%)
WEEK 84	35/96 (36.5%)	10/49 (20.4%)	45/145 (31.0%)
WEEK 90	28/96 (29.2%)	8/49 (16.3%)	36/145 (24.8%)
WEEK 96	29/96 (30.2%)	7/49 (14.3%)	36/145 (24.8%)
WEEK 102	18/96 (18.8%)	6/49 (12.2%)	24/145 (16.6%)
WEEK 108	18/96 (18.8%)	9/49 (18.4%)	27/145 (18.6%)
WEEK 114	11/96 (11.5%)	5/49 (10.2%)	16/145 (11.0%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least half of the 6 weekly item scores for that week (visit) non-missing (i.e., >=3 weekly items).  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 5

Table 14.1.1.1  
 Summary of Completion Rate of NTD-T-PRO  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 120	12/96 (12.5%)	0/49 (0.0%)	12/145 (8.3%)
WEEK 126	6/96 (6.3%)	0/49 (0.0%)	6/145 (4.1%)
WEEK 132	2/96 (2.1%)	0/49 (0.0%)	2/145 (1.4%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least half of the 6 weekly item scores for that week (visit) non-missing (i.e., >=3 weekly items).  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Tiredness/weakness (T/W)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	4.1 (2.31)	4.0 (2.03)	
Median	4.4	4.1	
Min, Max	0.0, 9.5	0.4, 9.5	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Tiredness/weakness (T/W)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	3.0 (2.18)	3.5 (1.80)	
Median	3.0	3.5	
Min, Max	0.0, 8.5	0.5, 7.5	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Tiredness/weakness (T/W)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	3.0 (2.23)	3.5 (2.03)	
Median	2.7	3.5	
Min, Max	0.0, 9.0	0.5, 8.8	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Tiredness/weakness (T/W)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-1.0 (-1.38, -0.64)	-0.4 (-0.89, 0.13)	-0.6 (-1.26, 0.00) 0.0494
			-0.4 (-0.77, 0.00)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Tiredness/weakness (T/W)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.9 (-1.30, -0.54)	-0.3 (-0.78, 0.26)	-0.7 (-1.31, -0.02) 0.0433 -0.5 (-0.96, 0.00)
Overall	-0.9 (-1.25, -0.54)	-0.4 (-0.86, 0.11)	-0.5 (-1.12, 0.07) 0.0854 -0.3 (-0.65, 0.05)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Shortness of Breath (SoB)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	3.4 (2.40)	3.2 (2.16)	
Median	3.5	3.4	
Min, Max	0.0, 9.0	0.0, 9.4	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms.sas

18MAY2022 13:38

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Shortness of Breath (SoB)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	2.6 (2.29)	3.0 (1.94)	
Median	2.2	3.0	
Min, Max	0.0, 9.0	0.0, 7.6	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Shortness of Breath (SoB)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	2.4 (2.17)	3.1 (2.43)	
Median	2.1	3.0	
Min, Max	0.0, 8.5	0.0, 9.2	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Shortness of Breath (SoB)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-0.7 (-1.08, -0.28)	0.0 (-0.55, 0.53)	-0.7 (-1.30, -0.05) 0.0356 -0.4 (-0.80, -0.02)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Shortness of Breath (SoB)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.7 (-1.14, -0.33)	0.2 (-0.37, 0.72)	-0.9 (-1.55, -0.27) 0.0054 -0.7 (-1.15, -0.18)
Overall	-0.7 (-1.04, -0.27)	-0.1 (-0.61, 0.43)	-0.6 (-1.16, 0.04) 0.0677 -0.3 (-0.67, 0.03)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 1 - Tiredness when not doing physical activity (TiredNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	3.3 (2.26)	3.1 (2.09)	
Median	3.1	3.0	
Min, Max	0.0, 9.0	0.0, 9.0	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 1 - Tiredness when not doing physical activity (TiredNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	2.3 (2.01)	2.4 (1.82)	
Median	2.0	2.0	
Min, Max	0.0, 7.0	0.0, 7.6	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 1 - Tiredness when not doing physical activity (TiredNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	2.4 (2.29)	2.4 (2.07)	
Median	2.0	2.0	
Min, Max	0.0, 10.0	0.0, 7.8	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 1 - Tiredness when not doing physical activity (TiredNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-0.9 (-1.33, -0.55)	-0.6 (-1.17, -0.10)	-0.3 (-0.91, 0.32) 0.3433 -0.2 (-0.57, 0.20)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 1 - Tiredness when not doing physical activity (TiredNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.7 (-1.13, -0.32)	-0.5 (-1.07, 0.03)	-0.2 (-0.84, 0.43) 0.5277 -0.1 (-0.62, 0.32)
Overall	-0.8 (-1.15, -0.41)	-0.6 (-1.09, -0.08)	-0.2 (-0.78, 0.37) 0.4911 -0.1 (-0.47, 0.23)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 2 - Tiredness when doing physical activity (TiredPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	5.0 (2.57)	5.0 (2.28)	
Median	5.3	5.2	
Min, Max	0.0, 10.0	0.3, 9.9	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 17 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 2 - Tiredness when doing physical activity (TiredPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	3.8 (2.62)	4.6 (2.15)	
Median	4.0	4.2	
Min, Max	0.0, 10.0	1.0, 8.1	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 18 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 2 - Tiredness when doing physical activity (TiredPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	3.8 (2.60)	4.3 (2.26)	
Median	3.6	4.1	
Min, Max	0.0, 10.0	1.0, 8.8	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 19 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 2 - Tiredness when doing physical activity (TiredPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-1.2 (-1.65, -0.69)	-0.3 (-0.98, 0.30)	-0.8 (-1.56, -0.10) 0.0267 -0.4 (-0.82, -0.04)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 20 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 2 - Tiredness when doing physical activity (TiredPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-1.1 (-1.61, -0.62)	-0.3 (-0.99, 0.32)	-0.8 (-1.53, -0.03) 0.0428 -0.5 (-0.96, 0.00)
Overall	-1.1 (-1.53, -0.61)	-0.4 (-1.03, 0.20)	-0.7 (-1.35, 0.04) 0.0638 -0.3 (-0.67, 0.02)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 21 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 3 - Weakness when not doing physical activity (WeakNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	3.2 (2.29)	3.0 (2.07)	
Median	3.0	3.0	
Min, Max	0.0, 9.0	0.0, 9.3	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 22 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 3 - Weakness when not doing physical activity (WeakNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	2.3 (1.98)	2.4 (1.83)	
Median	2.0	2.3	
Min, Max	0.0, 7.0	0.0, 7.2	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 23 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 3 - Weakness when not doing physical activity (WeakNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Week 48			
N [a]	48	27	
Mean (SD)	2.2 (2.06)	2.6 (2.32)	
Median	1.7	2.0	
Min, Max	0.0, 8.0	0.0, 9.5	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 24 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 3 - Weakness when not doing physical activity (WeakNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-0.9 (-1.30, -0.52)	-0.5 (-1.03, 0.04)	-0.4 (-1.03, 0.21) 0.1921 -0.3 (-0.64, 0.13)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 25 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 3 - Weakness when not doing physical activity (WeakNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.8 (-1.20, -0.39)	-0.2 (-0.77, 0.32)	-0.6 (-1.21, 0.07) 0.0790 -0.4 (-0.89, 0.06)
Overall	-0.8 (-1.18, -0.43)	-0.4 (-0.93, 0.10)	-0.4 (-0.98, 0.19) 0.1888 -0.2 (-0.58, 0.12)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 26 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 4 - Weakness when doing physical activity (WeakPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	4.9 (2.68)	4.9 (2.35)	
Median	5.0	5.0	
Min, Max	0.0, 10.0	0.4, 10.0	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 27 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 4 - Weakness when doing physical activity (WeakPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	3.7 (2.68)	4.7 (2.16)	
Median	3.7	4.0	
Min, Max	0.0, 10.0	1.0, 8.3	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 28 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 4 - Weakness when doing physical activity (WeakPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	3.6 (2.56)	4.5 (2.24)	
Median	3.2	4.5	
Min, Max	0.0, 10.0	1.0, 9.0	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 29 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 4 - Weakness when doing physical activity (WeakPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-1.1 (-1.63, -0.63)	-0.2 (-0.86, 0.46)	-0.9 (-1.68, -0.18) 0.0157 -0.5 (-0.86, -0.08)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 30 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 4 - Weakness when doing physical activity (WeakPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-1.2 (-1.69, -0.67)	-0.1 (-0.76, 0.57)	-1.1 (-1.86, -0.32) 0.0057 -0.7 (-1.14, -0.18)
Overall	-1.0 (-1.52, -0.55)	-0.2 (-0.87, 0.40)	-0.8 (-1.53, -0.09) 0.0281 -0.4 (-0.74, -0.04)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 31 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 5 - Shortness of breath when not doing physical activity (SobNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	2.5 (2.19)	2.3 (2.00)	
Median	2.2	2.4	
Min, Max	0.0, 8.4	0.0, 8.9	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 32 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 5 - Shortness of breath when not doing physical activity (SobNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	1.9 (2.04)	2.1 (1.98)	
Median	1.5	2.0	
Min, Max	0.0, 8.0	0.0, 7.6	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 33 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 5 - Shortness of breath when not doing physical activity (SobNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	1.7 (1.96)	2.3 (2.48)	
Median	1.2	2.0	
Min, Max	0.0, 7.0	0.0, 8.8	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 34 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 5 - Shortness of breath when not doing physical activity (SobNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-0.5 (-0.89, -0.16)	-0.1 (-0.57, 0.43)	-0.5 (-1.04, 0.13) 0.1236 -0.3 (-0.69, 0.09)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 35 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 5 - Shortness of breath when not doing physical activity (SobNA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.5 (-0.92, -0.17)	0.1 (-0.37, 0.64)	-0.7 (-1.27, -0.08) 0.0257 -0.5 (-1.01, -0.05)
Overall	-0.5 (-0.85, -0.15)	-0.1 (-0.60, 0.36)	-0.4 (-0.93, 0.18) 0.1802 -0.2 (-0.58, 0.11)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 36 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 6 - Shortness of breath when doing physical activity (SobPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	4.2 (2.84)	4.1 (2.57)	
Median	4.5	4.2	
Min, Max	0.0, 10.0	0.0, 9.9	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 37 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDI-PRO by Treatment Group and Visit  
 (ITT Population)

Item 6 - Shortness of breath when doing physical activity (SobPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	3.3 (2.80)	4.0 (2.27)	
Median	2.8	4.0	
Min, Max	0.0, 10.0	0.0, 8.0	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDI-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 38 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 6 - Shortness of breath when doing physical activity (SobPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	3.2 (2.65)	3.9 (2.68)	
Median	2.6	4.0	
Min, Max	0.0, 10.0	0.0, 9.5	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 39 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTD-T-PRO by Treatment Group and Visit  
 (ITT Population)

Item 6 - Shortness of breath when doing physical activity (SobPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-0.8 (-1.33, -0.34)	0.1 (-0.62, 0.73)	-0.9 (-1.67, -0.11) 0.0258 -0.4 (-0.83, -0.05)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 40 of 40

Table 14.4.1  
 Observed Scores and Least Square Mean Changes from Baseline of NTDT-PRO by Treatment Group and Visit  
 (ITT Population)

Item 6 - Shortness of breath when doing physical activity (SobPA)

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.9 (-1.43, -0.43)	0.2 (-0.47, 0.89)	-1.1 (-1.93, -0.34) 0.0051 -0.7 (-1.15, -0.19)
Overall	-0.8 (-1.29, -0.33)	-0.1 (-0.71, 0.59)	-0.7 (-1.50, 0.01) 0.0522 -0.3 (-0.69, 0.01)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 4

Table 14.3.1.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event	KME (95%CI) (mon)	[a]	Subjects with event	KME (95%CI) (mon)	[a]	HR (95% CI) [b]	p-value [c]
Item 1 - Tiredness when not doing physical activity (TiredNA) Responder Definition=1.5 94	44 (46.8)	NE (4.4, NE)	49	21 (42.9)	NE (1.6, NE)	0.97 (0.57, 1.64)	0.9998	
Item 2 - Tiredness when doing physical activity (TiredPA) Responder Definition=1.5 94	51 (54.3)	4.7 (2.1, NE)	49	24 (49.0)	18.9 (1.6, NE)	1.19 (0.73, 1.96)	0.3267	

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category (>=3 points vs. <3 points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e. >=8.5 g/dL, <8.5 g/dL) and NTDT-PRO T/W domain score level (i.e. <3, >=3).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 4

Table 14.3.1.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	HR (95% CI) [b]	p-value [c]
Item 3 - Weakness when not doing physical activity (WeakNA) Responder Definition=1.5	94	41 (43.6)	NE (5.4, NE)	49	19 (38.8)	NE (2.8, NE)	1.04 (0.60, 1.81)	0.6684
Item 4 - Weakness when doing physical activity (WeakPA) Responder Definition=1.5	94	50 (53.2)	7.7 (1.9, NE)	49	25 (51.0)	18.9 (1.4, NE)	1.15 (0.70, 1.87)	0.5398

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category ( $\geq 3$  points vs.  $< 3$  points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e.  $\geq 8.5$  g/dL,  $< 8.5$  g/dL) and NTDT-PRO T/W domain score level (i.e.  $< 3$ ,  $\geq 3$ ).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 4

Table 14.3.1.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	HR (95% CI) [b]	p-value [c]
Item 5 - Shortness of breath when not doing physical activity (SobNA) Responder Definition=1.5	94	33 (35.1)	NE (23.1, NE)	49	14 (28.6)	NE (NE, NE)	1.13 (0.60, 2.13)	0.4127
Item 6 - Shortness of breath when doing physical activity (SobPA) Responder Definition=1.5	94	45 (47.9)	NE (2.8, NE)	49	22 (44.9)	NE (1.6, NE)	1.04 (0.62, 1.76)	0.4175

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category ( $\geq 3$  points vs.  $< 3$  points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e.  $\geq 8.5$  g/dL,  $< 8.5$  g/dL) and NTDT-PRO T/W domain score level (i.e.  $< 3$ ,  $\geq 3$ ).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 4

Table 14.3.1.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event	KME (95%CI) (mon)	[a]	Subjects with event	KME (95%CI) (mon)	[a]	HR (95% CI) [b]	p-value [c]
Tiredness/weakness (T/W) Responder Definition=1.5	94	45 (47.9)	NE (3.5, NE)	49	22 (44.9)	20.3 (1.6, NE)	1.03 (0.61, 1.71)	0.8972
Shortness of Breath (SoB) Responder Definition=1.5	94	40 (42.6)	NE (7.7, NE)	49	17 (34.7)	NE (11.9, NE)	1.20 (0.67, 2.14)	0.2348

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category (>=3 points vs. <3 points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e. >=8.5 g/dL, <8.5 g/dL) and NTDT-PRO T/W domain score level (i.e. <3, >=3).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 4

Table 14.3.1.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	HR (95% CI) [b]	p-value [c]
Item 1 - Tiredness when not doing physical activity (TiredNA) Responder Definition=1.5 94	20 (21.3)	NE (NE, NE)	49	9 (18.4)	NE (NE, NE)	1.21 (0.55, 2.66)	0.7221	
Item 2 - Tiredness when doing physical activity (TiredPA) Responder Definition=1.5 94	21 (22.3)	NE (NE, NE)	49	14 (28.6)	NE (NE, NE)	0.73 (0.37, 1.43)	0.3649	

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category ( $\geq 3$  points vs.  $< 3$  points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e.  $\geq 8.5$  g/dL,  $< 8.5$  g/dL) and NTDT-PRO T/W domain score level (i.e.  $< 3$ ,  $\geq 3$ ).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 4

Table 14.3.1.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	Subjects with event N	KME (95%CI) (mon) n (%)	[a]	HR (95% CI) [b]	p-value [c]
Item 3 - Weakness when not doing physical activity (WeakNA) Responder Definition=1.5	94	23 (24.5)	NE (NE, NE)	49	10 (20.4)	NE (NE, NE)	1.28 (0.61, 2.72)	0.6008
Item 4 - Weakness when doing physical activity (WeakPA) Responder Definition=1.5	94	23 (24.5)	NE (NE, NE)	49	15 (30.6)	NE (NE, NE)	0.70 (0.36, 1.36)	0.4216

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category ( $\geq 3$  points vs.  $< 3$  points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e.  $\geq 8.5$  g/dL,  $< 8.5$  g/dL) and NTDT-PRO T/W domain score level (i.e.  $< 3$ ,  $\geq 3$ ).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 4

Table 14.3.1.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event	KME (95%CI) (mon)		Subjects with event	KME (95%CI) (mon)		HR (95% CI) [b]	p-value [c]
	N	n (%)	[a]	N	n (%)	[a]		
Item 5 - Shortness of breath when not doing physical activity (SobNA) Responder Definition=1.5	94	25 (26.6)	NE (NE, NE)	49	11 (22.4)	NE (NE, NE)	1.29 (0.63, 2.62)	0.5319
Item 6 - Shortness of breath when doing physical activity (SobPA) Responder Definition=1.5	94	26 (27.7)	NE (30.1, NE)	49	15 (30.6)	NE (NE, NE)	0.82 (0.43, 1.57)	0.5513

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category (>=3 points vs. <3 points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e. >=8.5 g/dL, <8.5 g/dL) and NTDT-PRO T/W domain score level (i.e. <3, >=3).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 4

Table 14.3.1.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event	KME (95%CI) (mon)	[a]	Subjects with event	KME (95%CI) (mon)	[a]	HR (95% CI) [b]	p-value [c]
Tiredness/weakness (T/W) Responder Definition=1.5	94	19 (20.2)	NE (NE, NE)	49	10 (20.4)	NE (NE, NE)	0.94 (0.43, 2.04)	0.8670
Shortness of Breath (SoB) Responder Definition=1.5	94	23 (24.5)	NE (NE, NE)	49	12 (24.5)	NE (NE, NE)	1.00 (0.49, 2.01)	0.9714

Notes: The ITT population is defined as all randomized subjects. The model for NTDT-PRO T/W score will not be adjusted by the randomization stratification factor baseline NTDT-PRO T/W score category (>=3 points vs. <3 points) since the baseline score is included in the model as a continuous variable.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Stratified Log-rank test.

Stratified by hemoglobin level (i.e. >=8.5 g/dL, <8.5 g/dL) and NTDT-PRO T/W domain score level (i.e. <3, >=3).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	94	45 (47.9)	NE (3.5, NE)	49	22 (44.9)	20.3 (1.6, NE)	1.05 (0.63, 1.75)	0.7593	
Region								0.4390	
North America & Europe	58	36 (62.1)	4.4 (1.2, 9.1)	30	15 (50.0)	13.3 (1.2, NE)	1.35 (0.73, 2.47)	0.2712	
Middle East	9	2 (22.2)	NE (0.9, NE)	8	2 (25.0)	NE (0.2, NE)	0.41 (0.05, 3.40)	0.8171	
Asia-Pacific	27	7 (25.9)	NE (NE, NE)	11	5 (45.5)	NE (0.2, NE)	0.52 (0.16, 1.68)	0.2745	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.2704
<=32	34	16 (47.1)	10.5 (4.4, NE)	12	6 (50.0)	20.3 (0.2, NE)	0.60 (0.22, 1.65)		
>32	60	29 (48.3)	NE (1.6, NE)	37	16 (43.2)	NE (1.2, NE)	1.17 (0.64, 2.16)		
Splenectomy									0.5557
Yes	33	16 (48.5)	NE (2.6, NE)	26	11 (42.3)	NE (1.6, NE)	1.21 (0.56, 2.63)		
No	61	29 (47.5)	NE (1.9, NE)	23	11 (47.8)	20.3 (0.7, NE)	0.86 (0.43, 1.73)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Sex								0.0662	
Male	39	17 (43.6)	NE (2.1, NE)	23	13 (56.5)	11.9 (0.7, NE)	0.64 (0.31, 1.34)	0.3483	
Female	55	28 (50.9)	9.1 (1.6, NE)	26	9 (34.6)	NE (1.6, NE)	1.67 (0.79, 3.55)	0.2032	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.2857
Beta-Thalassemia	61	37 (60.7)	4.4 (1.2, 10.5)	34	15 (44.1)	NE (1.2, NE)	1.36 (0.74, 2.50)	0.1377	
Hemoglobin E/Beta-Thalassemia	28	6 (21.4)	NE (NE, NE)	11	5 (45.5)	NE (0.2, NE)	0.39 (0.12, 1.34)	0.1588	
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (0.2, NE)	4	2 (50.0)	13.3 (0.7, NE)	1.08 (0.14, 8.08)	0.8214	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia gene mutation								0.7875
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	68	31 (45.6)	NE (4.4, NE)	33	12 (36.4)	NE (4.7, NE)	1.13 (0.57, 2.23)	0.4013
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	6	2 (33.3)	NE (0.2, NE)	4	2 (50.0)	13.3 (0.7, NE)	1.06 (0.14, 7.96)	0.7030
[B0/B, B+/B] with Alpha gene duplication	20	12 (60.0)	1.5 (0.9, NE)	12	8 (66.7)	1.2 (0.5, NE)	0.80 (0.32, 1.98)	0.6158

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	53	24 (45.3)	NE (3.5, NE)	29	8 (27.6)	NE (20.3, NE)	1.80 (0.80, 4.04)	0.1324	0.0354*
>= 8.5 g/dL	41	21 (51.2)	9.1 (0.9, NE)	20	14 (70.0)	1.2 (0.5, 11.9)	0.62 (0.32, 1.23)	0.1679	
Baseline NTDT-PRO T/W score <3 points	30	4 (13.3)	NE (NE, NE)	14	3 (21.4)	NE (20.3, NE)	1.04 (0.22, 4.84)	0.6235	0.5706
>=3 points	64	41 (64.1)	2.6 (0.9, 7.7)	35	19 (54.3)	4.7 (0.7, NE)	1.13 (0.66, 1.96)	0.5335	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.7242	
<3 mg/gr dry weight	35	17 (48.6)	9.1 (3.0, NE)	17	9 (52.9)	1.6 (0.5, NE)	0.75 (0.33, 1.69)	0.5884	
>=3-<=5 mg/gr dry weight	22	11 (50.0)	10.5 (0.9, NE)	13	5 (38.5)	20.3 (0.7, 20.3)	1.23 (0.42, 3.62)	0.6404	
>5-<=7 mg/gr dry weight	12	7 (58.3)	1.3 (0.5, NE)	3	1 (33.3)	NE (0.7, NE)	1.25 (0.14, 10.93)	0.6389	
>7-<=15 mg/gr dry weight	13	5 (38.5)	NE (0.7, NE)	9	4 (44.4)	NE (0.2, NE)	1.83 (0.44, 7.62)	0.9164	
>15 mg/gr dry weight	11	5 (45.5)	NE (0.5, NE)	5	3 (60.0)	11.9 (0.7, NE)	0.61 (0.14, 2.75)	0.8712	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline 6MWT <=450	50	23 (46.0)	NE (1.9, NE)	31	13 (41.9)	NE (0.9, NE)	0.93 (0.46, 1.88)	0.6651	
>450	44	22 (50.0)	9.1 (2.1, NE)	18	9 (50.0)	20.3 (1.2, NE)	1.18 (0.54, 2.60)	0.6742	
Baseline FACIT-F FS score <43 points	63	37 (58.7)	4.7 (1.6, NE)	32	18 (56.3)	1.6 (0.7, NE)	0.99 (0.56, 1.74)	0.7916	
>=43 points	31	8 (25.8)	NE (NE, NE)	16	4 (25.0)	NE (20.3, NE)	0.96 (0.28, 3.28)	0.9993	
							0.9222		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Overall	94	40 (42.6)	NE (7.7, NE)	49	17 (34.7)	NE (11.9, NE)	1.12 (0.63, 1.98) 0.4089	
Region								0.9148
North America & Europe	58	29 (50.0)	9.1 (3.0, NE)	30	14 (46.7)	NE (1.6, NE)	0.95 (0.50, 1.80) 0.9098	
Middle East	9	2 (22.2)	NE (0.9, NE)	8	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.1695	
Asia-Pacific	27	9 (33.3)	NE (3.0, NE)	11	3 (27.3)	NE (0.2, NE)	0.93 (0.24, 3.68) 0.7484	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.4904
<=32	34	14 (41.2)	NE (4.9, NE)	12	4 (33.3)	NE (0.2, NE)	0.84 (0.26, 2.68)		
>32	60	26 (43.3)	NE (3.0, NE)	37	13 (35.1)	NE (3.0, NE)	1.21 (0.62, 2.36)		
Splenectomy									0.1178
Yes	33	17 (51.5)	14.7 (2.6, NE)	26	7 (26.9)	NE (11.9, NE)	1.90 (0.78, 4.66)		
No	61	23 (37.7)	NE (7.7, NE)	23	10 (43.5)	NE (1.6, NE)	0.77 (0.37, 1.63)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Sex								0.8116	
Male	39	17 (43.6)	NE (2.8, NE)	23	8 (34.8)	NE (1.9, NE)	1.07 (0.45, 2.54)	0.5386	
Female	55	23 (41.8)	NE (4.9, NE)	26	9 (34.6)	NE (3.0, NE)	1.19 (0.55, 2.57)	0.5631	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.7506
Beta-Thalassemia	61	29 (47.5)	14.7 (3.5, NE)	34	13 (38.2)	NE (1.9, NE)	1.09 (0.56, 2.11)		
Hemoglobin E/Beta-Thalassemia	28	8 (28.6)	NE (NE, NE)	11	3 (27.3)	NE (0.2, NE)	0.83 (0.21, 3.34)		
Beta-Thalassemia combined with Alpha-Thalassemia	5	3 (60.0)	2.1 (0.9, NE)	4	1 (25.0)	NE (0.7, NE)	1.65 (0.16, 17.19)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia gene mutation								0.4896
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	68	28 (41.2)	NE (7.7, NE)	33	8 (24.2)	NE (NE, NE)	1.34 (0.59, 3.00)	0.1103
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	6	3 (50.0)	NE (0.9, NE)	4	1 (25.0)	NE (0.7, NE)	1.52 (0.15, 15.68)	0.7709
[B0/B, B+/B] with Alpha gene duplication	20	9 (45.0)	NE (1.2, NE)	12	8 (66.7)	2.3 (0.5, NE)	0.74 (0.28, 1.97)	0.2439

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	53	24 (45.3)	NE (3.5, NE)	29	6 (20.7)	NE (NE, NE)	2.03 (0.82, 5.04)	0.0751	
>= 8.5 g/dL	41	16 (39.0)	NE (7.7, NE)	20	11 (55.0)	7.5 (0.7, NE)	0.70 (0.32, 1.52)	0.0497	
Baseline NTD-T-PRO T/W score <3 points	30	4 (13.3)	NE (NE, NE)	14	3 (21.4)	NE (11.9, NE)	1.17 (0.22, 6.30)	0.5651	
>=3 points	64	36 (56.3)	4.9 (1.9, NE)	35	14 (40.0)	NE (1.9, NE)	1.25 (0.67, 2.34)	0.3092	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo			
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value	
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d]	[e]
Baseline liver iron										0.9907
<3 mg/gr dry weight	35	15 (42.9)	NE (3.0, NE)	17	9 (52.9)	3.0 (0.5, NE)	0.62	(0.27, 1.43)		
>=3-<=5 mg/gr dry weight	22	9 (40.9)	NE (1.2, NE)	13	5 (38.5)	NE (0.5, NE)	0.84	(0.28, 2.56)		
>5-<=7 mg/gr dry weight	12	8 (66.7)	6.3 (0.5, NE)	3	0 (0.0)	NE (NE, NE)	>9999	(<0.01, NE)		
>7-<=15 mg/gr dry weight	13	3 (23.1)	NE (14.7, NE)	9	0 (0.0)	NE (NE, NE)	>9999	(<0.01, NE)		
>15 mg/gr dry weight	11	5 (45.5)	NE (0.5, NE)	5	3 (60.0)	11.9 (0.7, NE)	0.60	(0.13, 2.69)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 16

Table 14.3.3.1.1  
 Summary of Time to First Improvement from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value	Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b][c]	[d][e]
Baseline 6MWT								0.3678
<=450	50	22 (44.0)	NE (3.0, NE)	31	11 (35.5)	NE (4.4, NE)	0.89 (0.42, 1.90)	0.5026
>450	44	18 (40.9)	NE (7.7, NE)	18	6 (33.3)	NE (1.6, NE)	1.53 (0.60, 3.91)	0.6627
Baseline FACIT-F FS score								0.4695
<43 points	63	34 (54.0)	7.7 (3.0, NE)	32	15 (46.9)	NE (1.2, NE)	1.01 (0.55, 1.86)	0.7562
>=43 points	31	6 (19.4)	NE (NE, NE)	16	2 (12.5)	NE (NE, NE)	1.84 (0.36, 9.26)	0.4731

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	94	19 (20.2)	NE (NE, NE)	49	10 (20.4)	NE (NE, NE)	0.94 (0.43, 2.03)	0.8372	
Region								0.0624	
North America & Europe	58	4 (6.9)	NE (NE, NE)	30	5 (16.7)	NE (NE, NE)	0.29 (0.07, 1.22)	0.0500	
Middle East	9	5 (55.6)	1.6 (0.5, NE)	8	2 (25.0)	NE (0.7, NE)	3.34 (0.64, 17.55)	0.1524	
Asia-Pacific	27	10 (37.0)	NE (3.7, NE)	11	3 (27.3)	NE (3.5, NE)	1.48 (0.40, 5.46)	0.4881	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.4683
<=32	34	7 (20.6)	NE (27.6, NE)	12	4 (33.3)	NE (2.3, NE)	0.59 (0.17, 2.10)		
>32	60	12 (20.0)	NE (NE, NE)	37	6 (16.2)	NE (20.3, NE)	1.20 (0.45, 3.21)		
Splenectomy									0.1171
Yes	33	8 (24.2)	NE (NE, NE)	26	3 (11.5)	NE (20.3, NE)	2.08 (0.55, 7.86)		
No	61	11 (18.0)	NE (27.6, NE)	23	7 (30.4)	NE (11.9, NE)	0.55 (0.21, 1.46)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Sex									0.2264
Male	39	9 (23.1)	NE (NE, NE)	23	3 (13.0)	NE (20.3, NE)	1.81 (0.49, 6.71)		0.4071
Female	55	10 (18.2)	NE (27.6, NE)	26	7 (26.9)	NE (11.9, NE)	0.60 (0.22, 1.61)		0.2889

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia diagnosis								0.7283
Beta-Thalassemia	61	8 (13.1)	NE (NE, NE)	34	7 (20.6)	NE (NE, NE)	0.67 (0.23, 1.94)	0.2401
Hemoglobin	28	10 (35.7)	NE (3.7, NE)	11	3 (27.3)	NE (3.5, NE)	1.42 (0.38, 5.25)	0.5294
E/Beta-Thalassemia								
Beta-Thalassemia combined with Alpha-Thalassemia	5	1 (20.0)	NE (0.7, NE)	4	0 (0.0)	NE (NE, NE)	<0.01 (<0.01, NE)	0.3711

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia gene mutation								0.2779	
[B0/B0, B+/B+, B+/B0]	68	17 (25.0)	NE (27.6, NE)	33	7 (21.2)	NE (20.3, NE)	1.20 (0.49, 2.93)	0.7273	
without Alpha-Thalassemia									
[B0/B0, B+/B+, B+/B0]	6	1 (16.7)	NE (0.7, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.4142	
with Alpha-Thalassemia									
[B0/B, B+/B] with Alpha gene duplication	20	1 (5.0)	NE (NE, NE)	12	3 (25.0)	NE (0.9, NE)	0.21 (0.02, 2.05)	0.1171	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	53	13 (24.5)	NE (27.6, NE)	29	5 (17.2)	NE (20.3, NE)	1.24 (0.44, 3.54)	0.6308	0.3930
>= 8.5 g/dL	41	6 (14.6)	NE (NE, NE)	20	5 (25.0)	NE (9.1, NE)	0.65 (0.20, 2.14)	0.3841	
Baseline NTDT-PRO T/W score <3 points	30	12 (40.0)	27.6 (3.5, NE)	14	5 (35.7)	NE (3.5, NE)	1.27 (0.43, 3.70)	0.6942	0.5559
>=3 points	64	7 (10.9)	NE (NE, NE)	35	5 (14.3)	NE (20.3, NE)	0.80 (0.25, 2.54)	0.4954	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.7355	
<3 mg/gr dry weight	35	6 (17.1)	NE (27.6, NE)	17	3 (17.6)	NE (11.9, NE)	0.79 (0.19, 3.37)	0.8076	
>=3-<=5 mg/gr dry weight	22	4 (18.2)	NE (NE, NE)	13	4 (30.8)	NE (0.7, NE)	0.44 (0.11, 1.79)	0.2683	
>5-<=7 mg/gr dry weight	12	1 (8.3)	NE (NE, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.6171	
>7-<=15 mg/gr dry weight	13	4 (30.8)	NE (1.6, NE)	9	1 (11.1)	NE (20.3, NE)	3.02 (0.31, 29.30)	0.2486	
>15 mg/gr dry weight	11	3 (27.3)	NE (1.9, NE)	5	1 (20.0)	NE (3.5, NE)	2.16 (0.21, 22.48)	0.7740	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Tiredness/weakness (T/W) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline 6MWT								0.0376*	
<=450	50	14 (28.0)	NE (27.6, NE)	31	6 (19.4)	NE (20.3, NE)	1.76 (0.67, 4.66)	0.4631	
>450	44	5 (11.4)	NE (NE, NE)	18	4 (22.2)	NE (2.3, NE)	0.39 (0.10, 1.50)	0.2656	
Baseline FACIT-F FS score								0.3269	
<43 points	63	12 (19.0)	NE (NE, NE)	32	5 (15.6)	NE (20.3, NE)	1.35 (0.47, 3.84)	0.7192	
>=43 points	31	7 (22.6)	NE (27.6, NE)	16	5 (31.3)	NE (3.5, NE)	0.61 (0.18, 2.00)	0.4455	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]	
Overall	94	23 (24.5)	NE (NE, NE)	49	12 (24.5)	NE (NE, NE)	1.04 (0.52, 2.10)	0.9978	
Region								0.5030	
North America & Europe	58	11 (19.0)	NE (NE, NE)	30	5 (16.7)	NE (NE, NE)	1.07 (0.37, 3.10)	0.9635	
Middle East	9	5 (55.6)	2.1 (0.7, NE)	8	3 (37.5)	NE (0.7, NE)	2.06 (0.48, 8.86)	0.3891	
Asia-Pacific	27	7 (25.9)	NE (NE, NE)	11	4 (36.4)	NE (0.5, NE)	0.71 (0.21, 2.41)	0.5556	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Age								0.3626	
<=32	34	7 (20.6)	NE (NE, NE)	12	4 (33.3)	NE (0.7, NE)	0.67 (0.19, 2.30)	0.3399	
>32	60	16 (26.7)	NE (NE, NE)	37	8 (21.6)	NE (NE, NE)	1.31 (0.56, 3.06)	0.5754	
Splenectomy								0.0321*	
Yes	33	10 (30.3)	NE (NE, NE)	26	3 (11.5)	NE (NE, NE)	2.92 (0.80, 10.65)	0.1049	
No	61	13 (21.3)	NE (NE, NE)	23	9 (39.1)	NE (3.5, NE)	0.52 (0.22, 1.23)	0.1072	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Sex									0.5300
Male	39	9 (23.1)	NE (NE, NE)	23	4 (17.4)	NE (NE, NE)	1.45 (0.45, 4.71)	0.5491	
Female	55	14 (25.5)	NE (NE, NE)	26	8 (30.8)	NE (11.9, NE)	0.89 (0.37, 2.12)	0.6102	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.7632
Beta-Thalassemia	61	14 (23.0)	NE (NE, NE)	34	8 (23.5)	NE (NE, NE)	1.13 (0.47, 2.72)		
Hemoglobin E/Beta-Thalassemia	28	7 (25.0)	NE (NE, NE)	11	4 (36.4)	NE (0.5, NE)	0.66 (0.19, 2.26)		
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (1.2, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia gene mutation								0.7963
[B0/B0, B+/B+, B+/B0]	68	17 (25.0)	NE (NE, NE)	33	9 (27.3)	NE (NE, NE)	1.09 (0.48, 2.45)	0.8988
without Alpha-Thalassemia								
[B0/B0, B+/B+, B+/B0]	6	2 (33.3)	NE (1.2, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2945
with Alpha-Thalassemia								
[B0/B, B+/B] with Alpha gene duplication	20	4 (20.0)	NE (NE, NE)	12	3 (25.0)	NE (2.6, NE)	0.58 (0.12, 2.72)	0.6515

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTD-T-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	53	11 (20.8)	NE (NE, NE)	29	5 (17.2)	NE (NE, NE)	1.30 (0.45, 3.76)	0.7119	0.5049
>= 8.5 g/dL	41	12 (29.3)	NE (15.6, NE)	20	7 (35.0)	NE (2.3, NE)	0.80 (0.31, 2.05)	0.6452	
Baseline NTD-T-PRO T/W score <3 points	30	10 (33.3)	NE (1.6, NE)	14	6 (42.9)	NE (2.1, NE)	0.89 (0.32, 2.47)	0.8407	0.6250
>=3 points	64	13 (20.3)	NE (NE, NE)	35	6 (17.1)	NE (NE, NE)	1.22 (0.46, 3.24)	0.8613	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTD-T-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline liver iron <3 mg/gr dry weight	35	10 (28.6)	NE (NE, NE)	17	4 (23.5)	NE (11.9, NE)	1.34 (0.42, 4.27)		0.9581
>=3-<=5 mg/gr dry weight	22	4 (18.2)	NE (NE, NE)	13	4 (30.8)	NE (2.6, NE)	0.60 (0.15, 2.41)		
>5-<=7 mg/gr dry weight	12	2 (16.7)	NE (0.2, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)		
>7-<=15 mg/gr dry weight	13	4 (30.8)	NE (1.2, NE)	9	2 (22.2)	NE (0.5, NE)	0.87 (0.13, 5.95)		
>15 mg/gr dry weight	11	2 (18.2)	NE (1.9, NE)	5	1 (20.0)	NE (3.5, NE)	1.27 (0.11, 15.34)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 16

Table 14.3.3.3.1  
 Summary of Time to First Deterioration from Baseline for NTDT-PRO - Subgroup Analysis  
 (ITT Population)

Shortness of Breath (SoB) (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Baseline 6MWT								0.6187
<=450	50	14 (28.0)	NE (NE, NE)	31	10 (32.3)	NE (9.1, NE)	1.10 (0.49, 2.48) 0.8514	
>450	44	9 (20.5)	NE (NE, NE)	18	2 (11.1)	NE (NE, NE)	1.64 (0.35, 7.60) 0.5010	
Baseline FACIT-F FS score								0.5515
<43 points	63	14 (22.2)	NE (NE, NE)	32	6 (18.8)	NE (NE, NE)	1.25 (0.48, 3.25) 0.7634	
>=43 points	31	9 (29.0)	NE (11.9, NE)	16	6 (37.5)	NE (3.5, NE)	0.83 (0.30, 2.35) 0.7308	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; NTDT-PRO = Non-transfusion-dependent Patient Reported Outcome.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 5

Table 14.1.1.4  
 Summary of Completion Rate of PGI-S  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
BASELINE	94/96 (97.9%)	49/49 (100.0%)	143/145 (98.6%)
WEEK 1	88/96 (91.7%)	45/49 (91.8%)	133/145 (91.7%)
WEEK 2	85/96 (88.5%)	45/49 (91.8%)	130/145 (89.7%)
WEEK 3	87/96 (90.6%)	48/49 (98.0%)	135/145 (93.1%)
WEEK 4	88/96 (91.7%)	45/49 (91.8%)	133/145 (91.7%)
WEEK 5	87/96 (90.6%)	46/49 (93.9%)	133/145 (91.7%)
WEEK 6	90/96 (93.8%)	45/49 (91.8%)	135/145 (93.1%)
WEEK 7	82/96 (85.4%)	43/49 (87.8%)	125/145 (86.2%)
WEEK 8	89/96 (92.7%)	46/49 (93.9%)	135/145 (93.1%)
WEEK 9	89/96 (92.7%)	44/49 (89.8%)	133/145 (91.7%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-S = Patient Global Impression of Severity.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as the weekly PGI-S for that week (visit) non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 5

Table 14.1.1.4  
 Summary of Completion Rate of PGI-S  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 10	87/96 (90.6%)	44/49 (89.8%)	131/145 (90.3%)
WEEK 11	89/96 (92.7%)	40/49 (81.6%)	129/145 (89.0%)
WEEK 12	85/96 (88.5%)	41/49 (83.7%)	126/145 (86.9%)
WEEK 13	87/96 (90.6%)	40/49 (81.6%)	127/145 (87.6%)
WEEK 14	85/96 (88.5%)	42/49 (85.7%)	127/145 (87.6%)
WEEK 15	81/96 (84.4%)	41/49 (83.7%)	122/145 (84.1%)
WEEK 16	82/96 (85.4%)	40/49 (81.6%)	122/145 (84.1%)
WEEK 17	82/96 (85.4%)	42/49 (85.7%)	124/145 (85.5%)
WEEK 18	85/96 (88.5%)	43/49 (87.8%)	128/145 (88.3%)
WEEK 19	82/96 (85.4%)	43/49 (87.8%)	125/145 (86.2%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-S = Patient Global Impression of Severity.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as the weekly PGI-S for that week (visit) non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 5

Table 14.1.1.4  
 Summary of Completion Rate of PGI-S  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 20	83/96 (86.5%)	41/49 (83.7%)	124/145 (85.5%)
WEEK 21	87/96 (90.6%)	38/49 (77.6%)	125/145 (86.2%)
WEEK 22	82/96 (85.4%)	40/49 (81.6%)	122/145 (84.1%)
WEEK 23	80/96 (83.3%)	36/49 (73.5%)	116/145 (80.0%)
WEEK 24	77/96 (80.2%)	39/49 (79.6%)	116/145 (80.0%)
WEEK 30	59/96 (61.5%)	31/49 (63.3%)	90/145 (62.1%)
WEEK 36	58/96 (60.4%)	33/49 (67.3%)	91/145 (62.8%)
WEEK 42	53/96 (55.2%)	23/49 (46.9%)	76/145 (52.4%)
WEEK 48	48/96 (50.0%)	27/49 (55.1%)	75/145 (51.7%)
WEEK 54	38/96 (39.6%)	16/49 (32.7%)	54/145 (37.2%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-S = Patient Global Impression of Severity.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as the weekly PGI-S for that week (visit) non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 5

Table 14.1.1.4  
 Summary of Completion Rate of PGI-S  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 60	39/96 (40.6%)	15/49 (30.6%)	54/145 (37.2%)
WEEK 66	39/96 (40.6%)	9/49 (18.4%)	48/145 (33.1%)
WEEK 72	41/96 (42.7%)	13/49 (26.5%)	54/145 (37.2%)
WEEK 78	35/96 (36.5%)	10/49 (20.4%)	45/145 (31.0%)
WEEK 84	35/96 (36.5%)	10/49 (20.4%)	45/145 (31.0%)
WEEK 90	28/96 (29.2%)	8/49 (16.3%)	36/145 (24.8%)
WEEK 96	29/96 (30.2%)	7/49 (14.3%)	36/145 (24.8%)
WEEK 102	18/96 (18.8%)	6/49 (12.2%)	24/145 (16.6%)
WEEK 108	18/96 (18.8%)	9/49 (18.4%)	27/145 (18.6%)
WEEK 114	11/96 (11.5%)	5/49 (10.2%)	16/145 (11.0%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-S = Patient Global Impression of Severity.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as the weekly PGI-S for that week (visit) non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 5

Table 14.1.1.4  
 Summary of Completion Rate of PGI-S  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 120	12/96 (12.5%)	0/49 (0.0%)	12/145 (8.3%)
WEEK 126	6/96 (6.3%)	0/49 (0.0%)	6/145 (4.1%)
WEEK 132	2/96 (2.1%)	0/49 (0.0%)	2/145 (1.4%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-S = Patient Global Impression of Severity.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as the weekly PGI-S for that week (visit) non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 3

Table 14.1.1.5  
 Summary of Completion Rate of PGI-C  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 6	74/96 (77.1%)	38/49 (77.6%)	112/145 (77.2%)
WEEK 12	77/96 (80.2%)	45/49 (91.8%)	122/145 (84.1%)
WEEK 18	75/96 (78.1%)	39/49 (79.6%)	114/145 (78.6%)
WEEK 24	75/96 (78.1%)	41/49 (83.7%)	116/145 (80.0%)
WEEK 30	67/96 (69.8%)	33/49 (67.3%)	100/145 (69.0%)
WEEK 36	78/96 (81.3%)	34/49 (69.4%)	112/145 (77.2%)
WEEK 42	68/96 (70.8%)	32/49 (65.3%)	100/145 (69.0%)
WEEK 48	73/96 (76.0%)	40/49 (81.6%)	113/145 (77.9%)
WEEK 54	61/96 (63.5%)	22/49 (44.9%)	83/145 (57.2%)
WEEK 60	61/96 (63.5%)	26/49 (53.1%)	87/145 (60.0%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-C = Patient Global Impression of Change.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as PGI-C for that visit non-missing.  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 3

Table 14.1.1.5  
 Summary of Completion Rate of PGI-C  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 66	60/96 (62.5%)	19/49 (38.8%)	79/145 (54.5%)
WEEK 72	54/96 (56.3%)	18/49 (36.7%)	72/145 (49.7%)
WEEK 78	51/96 (53.1%)	15/49 (30.6%)	66/145 (45.5%)
WEEK 84	54/96 (56.3%)	13/49 (26.5%)	67/145 (46.2%)
WEEK 90	50/96 (52.1%)	13/49 (26.5%)	63/145 (43.4%)
WEEK 96	43/96 (44.8%)	8/49 (16.3%)	51/145 (35.2%)
WEEK 102	34/96 (35.4%)	6/49 (12.2%)	40/145 (27.6%)
WEEK 108	30/96 (31.3%)	9/49 (18.4%)	39/145 (26.9%)
WEEK 114	26/96 (27.1%)	9/49 (18.4%)	35/145 (24.1%)
WEEK 120	17/96 (17.7%)	1/49 (2.0%)	18/145 (12.4%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-C = Patient Global Impression of Change.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as PGI-C for that visit non-missing.  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 3

Table 14.1.1.5  
 Summary of Completion Rate of PGI-C  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 126	6/96 (6.3%)	0/49 (0.0%)	6/145 (4.1%)
WEEK 132	2/96 (2.1%)	0/49 (0.0%)	2/145 (1.4%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; PGI-C = Patient Global Impression of Change.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as PGI-C for that visit non-missing.  
 Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 1

Table 14.3.1.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event	N	KME (95%CI) (mon) [a]	Subjects with event	N	KME (95%CI) (mon) [a]	HR (95% CI) [b]	p-value [c]
PGI-C								
Strong improvement	94	70 (74.5)	4.4 (3.9, 5.7)	48	9 (18.8)	NE (NE, NE)	6.44 (3.19, 12.97)	<.0001
Any improvement	94	86 (91.5)	2.2 (1.5, 2.8)	48	22 (45.8)	22.7 (4.4, NE)	4.19 (2.53, 6.95)	<.0001

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Stratified Log-rank test.  
 Stratified by hemoglobin level (i.e.  $\geq 8.5$  g/dL,  $< 8.5$  g/dL) and NTD-T-PRO T/W domain score level (i.e.  $< 3$ ,  $\geq 3$ ).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 1

Table 14.3.1.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C  
 (ITT Population)

Item/Total Score	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) [b]	p-value [c]
PGI-C								
Strong deterioration	94	6 (6.4)	NE (NE, NE)	48	5 (10.4)	NE (NE, NE)	0.51 (0.15, 1.71)	0.2686
Any deterioration	94	20 (21.3)	NE (28.2, NE)	48	13 (27.1)	NE (19.6, NE)	0.66 (0.32, 1.34)	0.2424

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Stratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Stratified Log-rank test.  
 Stratified by hemoglobin level (i.e.  $\geq 8.5$  g/dL,  $< 8.5$  g/dL) and NTD-T-PRO T/W domain score level (i.e.  $< 3$ ,  $\geq 3$ ).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte.sas

26MAY2022 12:49

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 5

Table 14.4.4  
 Observed Scores and Least Square Mean Changes from Baseline of PGI-S by Treatment Group and Visit  
 (ITT Population)

PGI-S

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	49	
Mean (SD)	3.8 (2.46)	3.7 (2.37)	
Median	4.0	3.8	
Min, Max	0.0, 8.8	0.0, 9.5	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; PGI-S = Patient Global Impression of Severity; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 5

Table 14.4.4  
 Observed Scores and Least Square Mean Changes from Baseline of PGI-S by Treatment Group and Visit  
 (ITT Population)

PGI-S

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	77	39	
Mean (SD)	3.0 (2.24)	3.8 (2.42)	
Median	3.0	4.0	
Min, Max	0.0, 9.6	0.0, 8.9	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; PGI-S = Patient Global Impression of Severity; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 5

Table 14.4.4  
 Observed Scores and Least Square Mean Changes from Baseline of PGI-S by Treatment Group and Visit  
 (ITT Population)

PGI-S

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	48	27	
Mean (SD)	2.9 (2.02)	3.6 (2.47)	
Median	2.7	3.3	
Min, Max	0.0, 8.3	0.0, 7.2	

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; PGI-S = Patient Global Impression of Severity; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 5

Table 14.4.4  
 Observed Scores and Least Square Mean Changes from Baseline of PGI-S by Treatment Group and Visit  
 (ITT Population)

PGI-S

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	93	49	
Week 24	-0.7 (-1.12, -0.24)	0.3 (-0.31, 0.89)	-1.0 (-1.66, -0.29) 0.0056 -0.5 (-0.94, -0.15)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; PGI-S = Patient Global Impression of Severity; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 5

Table 14.4.4  
 Observed Scores and Least Square Mean Changes from Baseline of PGI-S by Treatment Group and Visit  
 (ITT Population)

PGI-S

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.7 (-1.11, -0.21)	0.4 (-0.22, 0.98)	-1.0 (-1.74, -0.34) 0.0037 -0.7 (-1.18, -0.21)
Overall	-0.6 (-1.03, -0.17)	0.2 (-0.37, 0.79)	-0.8 (-1.47, -0.15) 0.0162 -0.4 (-0.77, -0.07)

Notes: The ITT population is defined as all randomized subjects. Overall LS mean changes from baseline and LS mean differences were estimated with a weight of 1/48 applied to each timepoint during the first 24 weeks (Week 1, 2, 3, ... 24) and a weight of 1/8 applied to each timepoint during the second 24 weeks (Week 30, 36, 42 and 48).

CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; PGI-S = Patient Global Impression of Severity; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. A first-order autoregressive covariance structure was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrms.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	94	39 (41.5)	NE (4.4, NE)	49	11 (22.4)	NE (NE, NE)	2.23 (1.13, 4.38)	0.0453	
Region								0.3313	
North America & Europe	58	31 (53.4)	4.4 (1.6, NE)	30	8 (26.7)	NE (NE, NE)	2.78 (1.25, 6.18)	0.0351	
Middle East	9	3 (33.3)	NE (0.9, NE)	8	1 (12.5)	NE (0.2, NE)	2.24 (0.23, 21.88)	0.3878	
Asia-Pacific	27	5 (18.5)	NE (NE, NE)	11	2 (18.2)	NE (0.9, NE)	0.43 (0.07, 2.72)	0.9598	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-S = Patient Global Impression of Severity.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Age									0.1453
<=32	34	13 (38.2)	NE (3.0, NE)	12	3 (25.0)	NE (0.5, NE)	0.76 (0.20, 2.87)		
>32	60	26 (43.3)	NE (2.6, NE)	37	8 (21.6)	NE (NE, NE)	2.55 (1.14, 5.71)		
Splenectomy									0.0853
Yes	33	13 (39.4)	NE (1.9, NE)	26	3 (11.5)	NE (NE, NE)	4.12 (1.15, 14.75)		
No	61	26 (42.6)	NE (3.0, NE)	23	8 (34.8)	NE (1.2, NE)	1.22 (0.55, 2.70)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Sex								0.6485	
Male	39	16 (41.0)	NE (1.6, NE)	23	6 (26.1)	NE (NE, NE)	1.66 (0.64, 4.27)	0.3597	
Female	55	23 (41.8)	NE (3.7, NE)	26	5 (19.2)	NE (NE, NE)	2.70 (1.02, 7.16)	0.0618	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia diagnosis								0.4443
Beta-Thalassemia	61	30 (49.2)	7.7 (1.9, NE)	34	9 (26.5)	NE (NE, NE)	2.43 (1.13, 5.20) 0.0496	
Hemoglobin	28	6 (21.4)	NE (NE, NE)	11	2 (18.2)	NE (0.9, NE)	0.55 (0.10, 3.17) 0.8883	
E/Beta-Thalassemia								
Beta-Thalassemia combined with Alpha-Thalassemia	5	3 (60.0)	1.6 (0.2, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.1109	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]	
Beta-thalassemia gene mutation								0.4605	
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	68	25 (36.8)	NE (4.4, NE)	33	6 (18.2)	NE (NE, NE)	2.67 (1.07, 6.63)	0.0810	
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	6	4 (66.7)	2.7 (0.2, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.1016	
[B0/B, B+/B] with Alpha gene duplication	20	10 (50.0)	7.7 (0.9, NE)	12	5 (41.7)	NE (0.5, NE)	1.10 (0.37, 3.28)	0.7659	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline hemoglobin < 8.5 g/dL	53	22 (41.5)	NE (3.7, NE)	29	3 (10.3)	NE (NE, NE)	4.68 (1.39, 15.79)	0.1261	
>= 8.5 g/dL	41	17 (41.5)	NE (1.9, NE)	20	8 (40.0)	NE (0.9, NE)	1.22 (0.51, 2.90)	0.0068 0.9885	
Baseline NTDT-PRO T/W score <3 points	30	1 (3.3)	NE (NE, NE)	14	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.9889	
>=3 points	64	38 (59.4)	2.8 (1.6, NE)	35	11 (31.4)	NE (1.4, NE)	2.15 (1.09, 4.24)	0.4795 0.0316	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.8467	
<3 mg/gr dry weight	35	17 (48.6)	NE (1.6, NE)	17	5 (29.4)	NE (0.9, NE)	2.72 (0.96, 7.73)	0.2486	
>=3-<=5 mg/gr dry weight	22	8 (36.4)	NE (1.9, NE)	13	3 (23.1)	NE (1.4, NE)	1.60 (0.41, 6.29)	0.4067	
>5-<=7 mg/gr dry weight	12	5 (41.7)	NE (0.7, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2744	
>7-<=15 mg/gr dry weight	13	5 (38.5)	NE (2.8, NE)	9	2 (22.2)	NE (0.2, NE)	1.78 (0.34, 9.30)	0.4955	
>15 mg/gr dry weight	11	4 (36.4)	NE (0.9, NE)	5	1 (20.0)	NE (0.7, NE)	0.61 (0.05, 7.15)	0.6178	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 8

Table 14.3.3.1.4  
 Summary of Time to First Improvement from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value	Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b][c]	[d][e]
Baseline 6MWT								0.7712
<=450	50	22 (44.0)	NE (2.6, NE)	31	6 (19.4)	NE (NE, NE)	1.98 (0.79, 4.95)	0.0390
>450	44	17 (38.6)	NE (3.0, NE)	18	5 (27.8)	NE (1.2, NE)	2.30 (0.81, 6.58)	0.5486
Baseline FACIT-F FS score								0.6698
<43 points	63	32 (50.8)	7.7 (2.6, NE)	32	10 (31.3)	NE (1.4, NE)	1.95 (0.95, 4.00)	0.1674
>=43 points	31	7 (22.6)	NE (NE, NE)	16	1 (6.3)	NE (NE, NE)	1.86 (0.20, 17.47)	0.1512

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-S = Patient Global Impression of Severity.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b] [c]	[d] [e]	
Overall	94	70 (74.5)	4.4 (3.9, 5.7)	48	9 (18.8)	NE (NE, NE)	6.47 (3.22, 13.01)	<.0001	
Region								0.0615	
North America & Europe	58	48 (82.8)	4.2 (3.0, 5.4)	29	3 (10.3)	NE (18.1, NE)	15.85 (4.91, 51.23)	<.0001	
Middle East	9	4 (44.4)	NE (4.4, NE)	8	2 (25.0)	NE (1.4, NE)	1.81 (0.33, 9.92)	0.4890	
Asia-Pacific	27	18 (66.7)	4.2 (2.8, NE)	11	4 (36.4)	NE (4.2, NE)	2.60 (0.88, 7.72)	0.0643	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.1475
<=32	34	26 (76.5)	4.4 (2.8, 8.3)	12	4 (33.3)	NE (2.7, NE)	3.35 (1.16, 9.68)		
>32	60	44 (73.3)	4.2 (3.1, 5.6)	36	5 (13.9)	NE (NE, NE)	8.92 (3.52, 22.62)		
									<.0001
Splenectomy									0.3551
Yes	33	21 (63.6)	5.8 (4.4, 14.2)	25	5 (20.0)	NE (18.1, NE)	4.81 (1.80, 12.84)		
No	61	49 (80.3)	3.9 (2.8, 4.7)	23	4 (17.4)	NE (18.2, NE)	7.98 (2.86, 22.23)		
									<.0001

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.0914
Male	39	28 (71.8)	5.4 (4.4, 7.0)	22	1 (4.5)	NE (18.1, NE)	26.90 (3.65, 198.50) <.0001	
Female	55	42 (76.4)	4.2 (2.9, 5.6)	26	8 (30.8)	NE (18.2, NE)	3.80 (1.77, 8.15) 0.0002	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia diagnosis								0.3116	
Beta-Thalassemia	62	48 (77.4)	4.4 (3.3, 5.7)	34	5 (14.7)	NE (18.1, NE)	9.13 (3.61, 23.08)	<.0001	
Hemoglobin E/Beta-Thalassemia	27	18 (66.7)	4.2 (2.8, 18.2)	11	4 (36.4)	NE (4.2, NE)	2.64 (0.89, 7.84)	0.0602	
Beta-Thalassemia combined with Alpha-Thalassemia	5	4 (80.0)	7.0 (1.5, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.0532	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value	Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b] [c]	[d] [e]
Beta-thalassemia gene mutation								0.2242
[B0/B0, B+/B+, B+/B0]	68	47 (69.1)	5.4 (4.3, 7.0)	33	8 (24.2)	NE (18.2, NE)	4.31 (2.03, 9.15)	<.0001
without Alpha-Thalassemia [B0/B0, B+/B+, B+/B0]	5	4 (80.0)	7.0 (1.5, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.0532
with Alpha-Thalassemia [B0/B, B+/B] with Alpha gene duplication	21	19 (90.5)	2.9 (1.7, 3.9)	12	1 (8.3)	NE (NE, NE)	23.63 (3.11, 179.80)	<.0001

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Baseline hemoglobin < 8.5 g/dL	54	38 (70.4)	4.6 (4.2, 5.9)	28	6 (21.4)	NE (18.1, NE)	5.41	(2.27, 12.87)	0.5073
>= 8.5 g/dL	40	32 (80.0)	3.9 (2.9, 7.0)	20	3 (15.0)	NE (NE, NE)	8.36	(2.55, 27.48)	<.0001
Baseline NTDT-PRO T/W score <3 points	28	25 (89.3)	4.2 (2.8, 4.4)	14	4 (28.6)	NE (18.1, NE)	6.03	(2.05, 17.72)	0.8634
>=3 points	66	45 (68.2)	5.4 (3.9, 8.3)	34	5 (14.7)	NE (NE, NE)	7.06	(2.79, 17.84)	<.0001

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.6153	
<3 mg/gr dry weight	35	27 (77.1)	4.2 (2.9, 7.0)	17	2 (11.8)	NE (NE, NE)	10.97 (2.59, 46.48)	<.0001	
>=3-<=5 mg/gr dry weight	22	18 (81.8)	4.5 (2.8, 11.2)	13	2 (15.4)	NE (7.1, NE)	8.15 (1.88, 35.32)	0.0008	
>5-<=7 mg/gr dry weight	12	9 (75.0)	4.7 (1.5, NE)	2	1 (50.0)	18.1 (NE, NE)	2.55 (0.32, 20.39)	0.3592	
>7-<=15 mg/gr dry weight	13	8 (61.5)	4.4 (1.4, NE)	9	1 (11.1)	NE (4.2, NE)	7.69 (0.96, 61.76)	0.0229	
>15 mg/gr dry weight	11	7 (63.6)	5.7 (2.8, NE)	5	2 (40.0)	NE (11.2, NE)	3.42 (0.67, 17.44)	0.1199	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline 6MWT <=450	50	33 (66.0)	5.8 (3.9, 11.0)	31	7 (22.6)	NE (18.2, NE)	4.36 (1.92, 9.91)	0.1863	
>450	44	37 (84.1)	4.2 (2.7, 5.4)	17	2 (11.8)	NE (NE, NE)	12.58 (3.01, 52.53)	0.0001	
Baseline FACIT-F FS score <43 points	64	45 (70.3)	4.7 (3.3, 7.0)	31	5 (16.1)	NE (18.1, NE)	6.75 (2.67, 17.08)	0.8223	
>=43 points	30	25 (83.3)	4.3 (2.8, 5.9)	16	4 (25.0)	NE (18.2, NE)	6.00 (2.06, 17.45)	<.0001	
							0.0002		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Overall	94	86 (91.5)	2.2 (1.5, 2.8)	48	22 (45.8)	22.7 (4.4, NE)	3.90	(2.40, 6.33)	<.0001
Region									0.0113*
North America & Europe	58	55 (94.8)	1.7 (1.5, 2.8)	29	10 (34.5)	NE (5.5, NE)	7.22	(3.56, 14.65)	<.0001
Middle East	9	5 (55.6)	7.0 (1.4, NE)	8	5 (62.5)	12.7 (1.3, NE)	0.86	(0.25, 3.00)	0.8104
Asia-Pacific	27	26 (96.3)	1.5 (1.4, 2.8)	11	7 (63.6)	9.8 (1.4, NE)	2.85	(1.18, 6.85)	0.0070

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.6372
<=32	34	32 (94.1)	1.8 (1.5, 2.8)	12	7 (58.3)	7.1 (2.7, NE)	3.35 (1.42, 7.87)		
>32	60	54 (90.0)	2.8 (1.4, 2.9)	36	15 (41.7)	22.7 (5.5, NE)	3.97 (2.21, 7.15)		
Splenectomy									0.6431
Yes	33	28 (84.8)	2.8 (1.5, 4.4)	25	11 (44.0)	22.7 (4.2, 22.7)	3.82 (1.82, 8.03)		
No	61	58 (95.1)	1.5 (1.4, 2.8)	23	11 (47.8)	26.6 (2.8, NE)	3.97 (2.04, 7.73)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.0054*
Male	39	35 (89.7)	2.8 (1.5, 3.6)	22	4 (18.2)	22.7 (22.7, NE)	16.04 (4.83, 53.22)	<.0001
Female	55	51 (92.7)	1.7 (1.4, 2.8)	26	18 (69.2)	3.6 (2.7, 26.6)	2.08 (1.20, 3.61)	0.0057

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia diagnosis								0.2275	
Beta-Thalassemia	62	57 (91.9)	2.2 (1.5, 2.9)	34	13 (38.2)	22.7 (4.4, NE)	5.19 (2.76, 9.76)	<.0001	
Hemoglobin E/Beta-Thalassemia	27	25 (92.6)	1.5 (1.4, 2.8)	11	7 (63.6)	9.8 (1.4, NE)	2.48 (1.05, 5.90)	0.0178	
Beta-Thalassemia combined with Alpha-Thalassemia	5	4 (80.0)	2.8 (1.4, NE)	3	2 (66.7)	5.5 (1.5, NE)	1.40 (0.25, 7.71)	0.6892	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo			
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value	
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d]	[e]
Beta-thalassemia gene mutation										0.1756
[B0/B0, B+/B+, B+/B0]	68	62 (91.2)	2.7 (1.5, 2.8)	33	17 (51.5)	22.7 (4.2, NE)	3.41	(1.96, 5.94)		
without Alpha-Thalassemia										<.0001
[B0/B0, B+/B+, B+/B0]	5	4 (80.0)	2.8 (1.4, NE)	3	2 (66.7)	5.5 (1.5, NE)	1.40	(0.25, 7.71)		
with Alpha-Thalassemia										0.6892
[B0/B, B+/B] with Alpha gene duplication	21	20 (95.2)	1.7 (1.4, 2.9)	12	3 (25.0)	NE (1.4, NE)	8.22	(2.32, 29.09)		0.0002

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline hemoglobin < 8.5 g/dL	54	49 (90.7)	1.5 (1.5, 2.8)	28	13 (46.4)	26.6 (4.2, NE)	3.97 (2.12, 7.43)	0.6731	
>= 8.5 g/dL	40	37 (92.5)	2.9 (1.5, 3.9)	20	9 (45.0)	22.7 (1.4, 22.7)	4.37 (1.97, 9.69)	0.0001	
Baseline NTDT-PRO T/W score <3 points	28	25 (89.3)	2.1 (1.4, 2.8)	14	8 (57.1)	6.3 (1.4, NE)	2.15 (0.96, 4.85)	0.2629	
>=3 points	66	61 (92.4)	2.3 (1.5, 2.8)	34	14 (41.2)	22.7 (5.5, 26.6)	5.79 (3.05, 10.97)	<.0001	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Baseline liver iron									0.8812
<3 mg/gr dry weight	35	32 (91.4)	2.6 (1.5, 2.8)	17	8 (47.1)	NE (1.4, NE)	3.28	(1.46, 7.37)	
>=3-<=5 mg/gr dry weight	22	22 (100.0)	2.8 (1.4, 4.2)	13	5 (38.5)	26.6 (4.2, 26.6)	8.69	(2.82, 26.78)	0.0020
>5-<=7 mg/gr dry weight	12	12 (100.0)	1.5 (1.4, 3.9)	2	1 (50.0)	NE (1.5, NE)	3.77	(0.48, 29.79)	<.0001
>7-<=15 mg/gr dry weight	13	10 (76.9)	2.8 (1.4, 4.4)	9	4 (44.4)	22.7 (1.5, 22.7)	2.89	(0.89, 9.38)	0.1682
>15 mg/gr dry weight	11	9 (81.8)	1.6 (1.4, 9.1)	5	3 (60.0)	5.5 (2.8, NE)	2.25	(0.60, 8.45)	0.0602
									0.1898

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 16

Table 14.3.3.1.5  
 Summary of Time to First Improvement from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any improvement)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b][c]	[d][e]	
Baseline 6MWT									0.0463*
<=450	50	45 (90.0)	2.8 (1.5, 2.9)	31	17 (54.8)	22.7 (2.8, 26.6)	2.61 (1.48, 4.62)	0.0004	
>450	44	41 (93.2)	1.6 (1.4, 2.8)	17	5 (29.4)	NE (4.4, NE)	8.61 (3.24, 22.87)	<.0001	
Baseline FACIT-F FS score									0.2329
<43 points	64	57 (89.1)	2.8 (1.5, 3.1)	31	15 (48.4)	22.7 (4.2, 22.7)	3.13 (1.75, 5.60)	<.0001	
>=43 points	30	29 (96.7)	1.6 (1.4, 2.8)	16	7 (43.8)	26.6 (2.8, NE)	6.20 (2.46, 15.64)	<.0001	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]	
Overall	94	25 (26.6)	NE (27.6, NE)	49	17 (34.7)	NE (14.7, NE)	0.71 (0.38, 1.33)	0.2490	
Region								0.2902	
North America & Europe	58	13 (22.4)	NE (27.6, NE)	30	10 (33.3)	20.3 (5.6, NE)	0.52 (0.22, 1.23)	0.1513	
Middle East	9	4 (44.4)	23.1 (0.2, NE)	8	2 (25.0)	NE (2.6, NE)	2.23 (0.41, 12.23)	0.3378	
Asia-Pacific	27	8 (29.6)	NE (3.7, NE)	11	5 (45.5)	14.7 (0.2, NE)	0.69 (0.22, 2.11)	0.4447	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-S = Patient Global Impression of Severity.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.6912
<=32	34	9 (26.5)	27.6 (23.1, NE)	12	4 (33.3)	NE (2.3, NE)	0.82 (0.25, 2.76)		0.5804
>32	60	16 (26.7)	NE (NE, NE)	37	13 (35.1)	NE (5.6, NE)	0.72 (0.34, 1.50)		0.3739
Splenectomy									0.3545
Yes	33	10 (30.3)	NE (23.1, NE)	26	8 (30.8)	NE (14.7, NE)	1.00 (0.39, 2.57)		0.8734
No	61	15 (24.6)	NE (27.6, NE)	23	9 (39.1)	NE (3.5, NE)	0.57 (0.25, 1.32)		0.1709

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Sex								0.3790	
Male	39	12 (30.8)	NE (23.1, NE)	23	7 (30.4)	20.3 (5.6, NE)	0.98 (0.38, 2.53)	0.9139	
Female	55	13 (23.6)	NE (27.6, NE)	26	10 (38.5)	NE (4.2, NE)	0.55 (0.24, 1.27)	0.1341	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-S = Patient Global Impression of Severity.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia diagnosis								0.9986	
Beta-Thalassemia	61	15 (24.6)	NE (27.6, NE)	34	12 (35.3)	NE (5.6, NE)	0.62 (0.29, 1.36)	0.1871	
Hemoglobin E/Beta-Thalassemia	28	8 (28.6)	NE (NE, NE)	11	5 (45.5)	14.7 (0.2, NE)	0.67 (0.22, 2.06)	0.4033	
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (0.2, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2132	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia gene mutation								0.9998	
[B0/B0, B+/B+, B+/B0]	68	18 (26.5)	NE (27.6, NE)	33	13 (39.4)	NE (5.4, NE)	0.64 (0.31, 1.32)	0.1731	
without Alpha-Thalassemia									
[B0/B0, B+/B+, B+/B0]	6	2 (33.3)	NE (0.2, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2605	
with Alpha-Thalassemia									
[B0/B, B+/B] with Alpha gene duplication	20	5 (25.0)	NE (7.7, NE)	12	4 (33.3)	NE (0.9, NE)	0.73 (0.20, 2.72)	0.6280	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	53	12 (22.6)	NE (27.6, NE)	29	11 (37.9)	NE (5.6, NE)	0.46 (0.20, 1.07)	0.0647	0.2146
>= 8.5 g/dL	41	13 (31.7)	NE (NE, NE)	20	6 (30.0)	NE (2.3, NE)	1.17 (0.44, 3.09)	0.8374	
Baseline NTDT-PRO T/W score <3 points	30	11 (36.7)	27.6 (7.7, NE)	14	7 (50.0)	20.3 (4.2, NE)	0.77 (0.29, 2.05)	0.5358	0.8578
>=3 points	64	14 (21.9)	NE (NE, NE)	35	10 (28.6)	NE (NE, NE)	0.77 (0.34, 1.73)	0.4181	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.5037	
<3 mg/gr dry weight	35	8 (22.9)	NE (27.6, NE)	17	5 (29.4)	NE (4.7, NE)	0.55 (0.17, 1.78)	0.5103	
>=3-<=5 mg/gr dry weight	22	6 (27.3)	NE (9.1, NE)	13	8 (61.5)	4.2 (0.7, NE)	0.35 (0.12, 1.03)	0.0368	
>5-<=7 mg/gr dry weight	12	2 (16.7)	NE (0.5, NE)	3	1 (33.3)	20.3 (NE, NE)	0.43 (0.04, 4.79)	0.4563	
>7-<=15 mg/gr dry weight	13	5 (38.5)	NE (0.5, NE)	9	2 (22.2)	NE (2.6, NE)	2.04 (0.39, 10.71)	0.3596	
>15 mg/gr dry weight	11	3 (27.3)	NE (1.6, NE)	5	1 (20.0)	NE (3.5, NE)	2.36 (0.22, 25.02)	0.7213	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-S = Patient Global Impression of Severity.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 8

Table 14.3.3.3.4  
 Summary of Time to First Deterioration from Baseline for PGI-S - Subgroup Analysis  
 (ITT Population)

PGI-S (Responder Definition=1.5)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Baseline 6MWT								0.8543
<=450	50	15 (30.0)	NE (27.6, NE)	31	13 (41.9)	NE (5.4, NE)	0.76 (0.35, 1.62) 0.1972	
>450	44	10 (22.7)	NE (NE, NE)	18	4 (22.2)	NE (4.7, NE)	1.07 (0.33, 3.46) 0.8848	
Baseline FACIT-F FS score								0.0117*
<43 points	63	19 (30.2)	NE (NE, NE)	32	7 (21.9)	NE (20.3, NE)	1.60 (0.67, 3.80) 0.3498	
>=43 points	31	6 (19.4)	NE (27.6, NE)	16	9 (56.3)	10.2 (0.5, NE)	0.21 (0.07, 0.65) 0.0014	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-S = Patient Global Impression of Severity.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	94	6 (6.4)	NE (NE, NE)	48	5 (10.4)	NE (NE, NE)	0.52 (0.16, 1.72)	0.2769	
Region								0.6961	
North America & Europe	58	3 (5.2)	NE (NE, NE)	29	4 (13.8)	NE (NE, NE)	0.23 (0.04, 1.25)	0.0632	
Middle East	9	1 (11.1)	NE (11.2, NE)	8	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.3496	
Asia-Pacific	27	2 (7.4)	NE (NE, NE)	11	1 (9.1)	NE (19.6, NE)	0.86 (0.08, 9.50)	0.9025	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Age								0.9940
<=32	34	2 (5.9)	NE (NE, NE)	12	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.4309	
>32	60	4 (6.7)	NE (NE, NE)	36	5 (13.9)	NE (NE, NE)	0.41 (0.11, 1.54) 0.1708	
Splenectomy								0.4054
Yes	33	4 (12.1)	NE (NE, NE)	25	3 (12.0)	NE (19.6, NE)	0.81 (0.18, 3.71) 0.7845	
No	61	2 (3.3)	NE (NE, NE)	23	2 (8.7)	NE (NE, NE)	0.36 (0.05, 2.57) 0.2894	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.5590
Male	39	4 (10.3)	NE (NE, NE)	22	4 (18.2)	NE (19.6, NE)	0.36 (0.08, 1.54) 0.1522	
Female	55	2 (3.6)	NE (NE, NE)	26	1 (3.8)	NE (NE, NE)	0.95 (0.09, 10.44) 0.9636	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia diagnosis								0.2897	
Beta-Thalassemia	62	1 (1.6)	NE (NE, NE)	34	4 (11.8)	NE (NE, NE)	0.09 (0.01, 0.90)	0.0132	
Hemoglobin	27	3 (11.1)	NE (NE, NE)	11	1 (9.1)	NE (19.6, NE)	1.31 (0.14, 12.62)	0.8135	
E/Beta-Thalassemia									
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (8.6, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2457	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia gene mutation								1.0000	
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	68	4 (5.9)	NE (NE, NE)	33	3 (9.1)	NE (NE, NE)	0.56 (0.12, 2.52)	0.4427	
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	5	2 (40.0)	NE (8.6, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2457	
[B0/B, B+/B] with Alpha gene duplication	21	0 (0.0)	NE (NE, NE)	12	2 (16.7)	NE (5.8, NE)	<0.01 (<0.01, NE)	0.0559	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	54	4 (7.4)	NE (NE, NE)	28	2 (7.1)	NE (19.6, NE)	0.81 (0.15, 4.53)	0.3456	
>= 8.5 g/dL	40	2 (5.0)	NE (NE, NE)	20	3 (15.0)	NE (NE, NE)	0.30 (0.05, 1.82)	0.8109	
Baseline NTDT-PRO T/W score <3 points	28	4 (14.3)	NE (NE, NE)	14	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.9931	
>=3 points	66	2 (3.0)	NE (NE, NE)	34	5 (14.7)	NE (19.6, NE)	0.17 (0.03, 0.88)	0.1582	
							0.0167		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Baseline liver iron									0.9965
<3 mg/gr dry weight	35	0 (0.0)	NE (NE, NE)	17	1 (5.9)	NE (NE, NE)	<0.01	(<0.01, NE)	
>=3-<=5 mg/gr dry weight	22	3 (13.6)	NE (NE, NE)	13	2 (15.4)	NE (5.8, NE)	0.81	(0.14, 4.86)	
>5-<=7 mg/gr dry weight	12	0 (0.0)	NE (NE, NE)	2	1 (50.0)	NE (3.0, NE)	<0.01	(<0.01, NE)	
>7-<=15 mg/gr dry weight	13	2 (15.4)	NE (22.4, NE)	9	1 (11.1)	NE (19.6, NE)	1.10	(0.10, 12.24)	
>15 mg/gr dry weight	11	1 (9.1)	NE (7.0, NE)	5	0 (0.0)	NE (NE, NE)	>9999	(<0.01, NE)	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Strong deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline 6MWT <=450	50	4 (8.0)	NE (NE, NE)	31	2 (6.5)	NE (NE, NE)	1.14 (0.21, 6.29)	0.8772	0.1574
>450	44	2 (4.5)	NE (NE, NE)	17	3 (17.6)	NE (NE, NE)	0.17 (0.03, 1.09)	0.0367	
Baseline FACIT-F FS score <43 points	64	6 (9.4)	NE (NE, NE)	31	5 (16.1)	NE (19.6, NE)	0.48 (0.14, 1.58)	0.2159	0.9999
>=43 points	30	0 (0.0)	NE (NE, NE)	16	0 (0.0)	NE (NE, NE)	1.00 (1.00, 1.00)	NE	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	94	20 (21.3)	NE (28.2, NE)	48	13 (27.1)	NE (19.6, NE)	0.66 (0.33, 1.35)	0.2560	
Region								0.1042	
North America & Europe	58	9 (15.5)	NE (NE, NE)	29	8 (27.6)	NE (NE, NE)	0.35 (0.12, 1.00)	0.0395	
Middle East	9	6 (66.7)	5.6 (1.4, NE)	8	4 (50.0)	NE (2.8, NE)	1.76 (0.49, 6.28)	0.3741	
Asia-Pacific	27	5 (18.5)	NE (NE, NE)	11	1 (9.1)	NE (19.6, NE)	2.24 (0.26, 19.16)	0.4504	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Age									0.0739
<=32	34	9 (26.5)	NE (22.4, NE)	12	1 (8.3)	NE (NE, NE)	3.03	(0.38, 24.01)	
>32	60	11 (18.3)	NE (28.2, NE)	36	12 (33.3)	NE (19.6, NE)	0.44	(0.19, 1.02)	
Splenectomy									0.7140
Yes	33	10 (30.3)	28.2 (22.4, NE)	25	8 (32.0)	NE (19.6, NE)	0.75	(0.28, 1.96)	
No	61	10 (16.4)	NE (NE, NE)	23	5 (21.7)	NE (NE, NE)	0.69	(0.23, 2.03)	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Sex									0.1393
Male	39	9 (23.1)	NE (NE, NE)	22	9 (40.9)	19.6 (6.3, NE)	0.37 (0.14, 0.99)		0.0391
Female	55	11 (20.0)	NE (28.2, NE)	26	4 (15.4)	NE (NE, NE)	1.19 (0.37, 3.80)		0.7663

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia diagnosis								0.2360
Beta-Thalassemia	62	12 (19.4)	NE (28.2, NE)	34	12 (35.3)	NE (8.6, NE)	0.38 (0.17, 0.90)	0.0216
Hemoglobin	27	6 (22.2)	NE (NE, NE)	11	1 (9.1)	NE (19.6, NE)	2.70 (0.32, 22.43)	0.3368
E/Beta-Thalassemia								
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (1.4, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2457

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia gene mutation								0.3251	
[B0/B0, B+/B+, B+/B0]	68	16 (23.5)	NE (NE, NE)	33	9 (27.3)	NE (19.6, NE)	0.78 (0.34, 1.77)	0.5497	
without Alpha-Thalassemia									
[B0/B0, B+/B+, B+/B0]	5	2 (40.0)	NE (1.4, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2457	
with Alpha-Thalassemia									
[B0/B, B+/B] with Alpha gene duplication	21	2 (9.5)	NE (28.2, NE)	12	4 (33.3)	NE (5.6, NE)	0.13 (0.01, 1.16)	0.0314	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Baseline hemoglobin < 8.5 g/dL	54	14 (25.9)	NE (28.2, NE)	28	7 (25.0)	NE (19.6, NE)	0.85	(0.34, 2.15)	0.3083
>= 8.5 g/dL	40	6 (15.0)	NE (NE, NE)	20	6 (30.0)	NE (7.1, NE)	0.45	(0.14, 1.42)	0.1776
Baseline NTDT-PRO T/W score <3 points	28	7 (25.0)	NE (22.4, NE)	14	2 (14.3)	NE (NE, NE)	1.77	(0.37, 8.56)	
>=3 points	66	13 (19.7)	NE (28.2, NE)	34	11 (32.4)	NE (19.6, NE)	0.48	(0.21, 1.09)	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.9107	
<3 mg/gr dry weight	35	4 (11.4)	NE (NE, NE)	17	3 (17.6)	NE (NE, NE)	0.51 (0.11, 2.35)	0.3822	
>=3-<=5 mg/gr dry weight	22	4 (18.2)	NE (NE, NE)	13	4 (30.8)	NE (5.8, NE)	0.62 (0.15, 2.47)	0.4884	
>5-<=7 mg/gr dry weight	12	2 (16.7)	28.2 (22.4, NE)	2	1 (50.0)	NE (3.0, NE)	<0.01 (<0.01, NE)	0.0143	
>7-<=15 mg/gr dry weight	13	5 (38.5)	NE (5.6, NE)	9	4 (44.4)	19.6 (2.8, NE)	0.74 (0.20, 2.78)	0.6538	
>15 mg/gr dry weight	11	4 (36.4)	NE (1.4, NE)	5	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.1443	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 PGI-C = Patient Global Impression of Change.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 16

Table 14.3.3.3.5  
 Summary of Time to First Deterioration from Baseline for PGI-C - Subgroup Analysis  
 (ITT Population)

PGI-C (Any deterioration)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Baseline 6MWT								0.0970
<=450	50	13 (26.0)	NE (28.2, NE)	31	7 (22.6)	NE (19.6, NE)	1.03 (0.40, 2.63) 0.9526	
>450	44	7 (15.9)	NE (NE, NE)	17	6 (35.3)	NE (5.8, NE)	0.33 (0.11, 1.01) 0.0411	
Baseline FACIT-F FS score								0.1461
<43 points	64	19 (29.7)	NE (22.4, NE)	31	10 (32.3)	NE (19.6, NE)	0.78 (0.36, 1.69) 0.5224	
>=43 points	30	1 (3.3)	NE (NE, NE)	16	3 (18.8)	NE (NE, NE)	0.18 (0.02, 1.71) 0.0918	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; PGI-C = Patient Global Impression of Change.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Protocol: ACE-536-B-THAL-002

Page 1 of 6

Subgroup Analysis of Change from Baseline in 6MWT at Week 48  
ITT Population

Subgroup	Luspatercept (N = 96)		Placebo (N = 49)		Luspatercept vs Placebo		
	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Difference in Mean Change (95% CI); P-value	SMD as Hedges'g (95% CI)	Test for Interaction P-value
	N (a)	(b)	N (a)	(b)	(b)	(b)	(c)
<b>REGION</b>							
NORTH AMERICA & EUROPE	52	461.70 ( 83.11) 18.92 (6.26, 31.58)	22	449.36 ( 85.62) -15.85 (-35.33, 3.63)	34.77 (11.52, 58.02) 0.0039	0.65 (0.20, 1.10)	0.0434
MIDDLE EAST	8	359.50 (170.47) 62.53 (37.28, 87.79)	8	392.00 ( 81.36) 73.34 (48.08, 98.60)	-10.81 (-46.68, 25.06) 0.5264	-0.27 (-1.23, 0.68)	
ASIA-PACIFIC	26	424.37 ( 60.40) -27.93 (-41.78, -14.07)	11	412.00 ( 39.64) -17.05 (-38.40, 4.29)	-10.88 (-36.38, 14.63) 0.3923	-0.29 (-1.00, 0.41)	
<b>AGE</b>							
<= 32	31	465.85 ( 82.09) 14.78 (-0.03, 29.59)	10	441.17 ( 55.43) -2.47 (-28.69, 23.74)	17.25 (-12.96, 47.47) 0.2549	0.37 (-0.29, 1.03)	0.5738
> 32	55	426.85 ( 95.95) 7.06 (-7.84, 21.95)	31	423.95 ( 83.53) -0.42 (-20.27, 19.43)	7.48 (-17.34, 32.29) 0.5507	0.12 (-0.29, 0.53)	

(a) Baseline is defined as the last value on or before D1D1.  
 (b) Estimates are based on an ANCOVA model with treatment group and baseline 6MWT value as continuous covariate.  
 Hedges g is derived from the estimate of the LS mean difference between Luspatercept and Placebo.  
 (c) The interaction is p-value is based on an ANCOVA model with treatment group, baseline 6MWT value, the subgroup factor and the treatment-group\*subgroup-factor interaction.

Data after the efficacy cutoff date of a subject will be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-chgsub-ubr700.sas

04AUG2022:03:54:36

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

Protocol: ACE-536-B-THAL-002

Page 2 of 6

Subgroup Analysis of Change from Baseline in 6MWT at Week 48  
ITT Population

Subgroup	Luspatercept (N = 96)		Placebo (N = 49)		Luspatercept vs Placebo		
	N (a)	Change from Baseline: Mean (SD) (95% CI)	N (a)	Change from Baseline: Mean (SD) (95% CI)	Difference in Mean Change (95% CI); P-value (b)	SMD as Hedges'g (95% CI) (b)	Test for Interaction P-value (c)
SPLENECTOMY							
YES	29	442.92 (80.86) 18.24 (0.43, 36.06)	20	414.32 (63.69) 8.27 (-13.25, 29.79)	9.97 (-18.21, 38.16) 0.4799	0.18 (-0.33, 0.69)	0.6844
NO	57	439.88 (98.76) 6.28 (-8.14, 20.69)	21	441.32 (87.94) -11.60 (-35.34, 12.14)	17.87 (-9.90, 45.65) 0.2038	0.31 (-0.18, 0.79)	
SEX							
MALE	36	449.61 (109.73) 15.96 (-1.03, 32.95)	18	430.46 (74.25) 0.26 (-23.80, 24.32)	15.70 (-13.81, 45.21) 0.2906	0.27 (-0.25, 0.78)	0.9841
FEMALE	50	434.64 (78.68) 6.79 (-7.04, 20.63)	23	426.34 (81.27) -4.80 (-25.21, 15.61)	11.60 (-13.08, 36.27) 0.3519	0.22 (-0.25, 0.68)	

(a) Baseline is defined as the last value on or before D1D1.  
 (b) Estimates are based on an ANCOVA model with treatment group and baseline 6MWT value as continuous covariate. Hedges g is derived from the estimate of the LS mean difference between Luspatercept and Placebo.  
 (c) The interaction is p-value is based on an ANCOVA model with treatment group, baseline 6MWT value, the subgroup factor and the treatment-group\*subgroup-factor interaction.  
 Data after the efficacy cutoff date of a subject will be used to evaluate this endpoint.  
 Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-chgsub-ubr700.sas 04AUG2022:03:54:36

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

Protocol: ACE-536-B-THAL-002

Page 3 of 6

Subgroup Analysis of Change from Baseline in 6MWT at Week 48  
ITT Population

Subgroup	Luspatercept (N = 96)		Placebo (N = 49)		Luspatercept vs Placebo		
	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Difference in Mean Change (95% CI); P-value (b)	SMD as Hedges'g (95% CI) (b)	Test for Interaction P-value (c)
BETA-THALASSEMIA DIAGNOSIS							0.1086
BETA-THALASSEMIA	55 450.48 (107.36)	27.10 (14.05, 40.16)	27 439.34 ( 88.26)	1.73 (-16.91, 20.37)	25.37 (2.61, 48.14) 0.0294	0.46 (0.04, 0.88)	
HEMOGLOBIN E/BETA-THALASSEMIA	26 423.60 ( 60.33)	-25.18 (-40.41, -9.95)	11 412.00 ( 39.64)	-17.00 (-40.46, 6.46)	-8.19 (-36.21, 19.84) 0.5567	-0.20 (-0.90, 0.50)	
BETA-THALASSEMIA GENE MUTATION							0.3492
[B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	63 430.97 ( 98.50)	9.41 (-3.67, 22.48)	27 416.67 ( 69.63)	0.09 (-19.91, 20.09)	9.32 (-14.60, 33.23) 0.4409	0.16 (-0.26, 0.58)	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	18 479.94 ( 73.18)	14.87 (-9.73, 39.46)	11 467.66 ( 88.14)	-15.17 (-46.66, 16.31)	30.04 (-9.97, 70.05) 0.1348	0.52 (-0.20, 1.24)	

(a) Baseline is defined as the last value on or before D1D1.

(b) Estimates are based on an ANCOVA model with treatment group and baseline 6MWT value as continuous covariate.

Hedges g is derived from the estimate of the LS mean difference between Luspatercept and Placebo.

(c) The interaction is p-value is based on an ANCOVA model with treatment group, baseline 6MWT value, the subgroup factor and the treatment-group\*subgroup-factor interaction.

Data after the efficacy cutoff date of a subject will be used to evaluate this endpoint.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-chgsub-ubr700.sas

04AUG2022:03:54:36

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

Protocol: ACE-536-B-THAL-002

Page 4 of 6

Subgroup Analysis of Change from Baseline in 6MWT at Week 48  
ITT Population

Subgroup	Luspatercept (N = 96)		Placebo (N = 49)		Luspatercept vs Placebo		
	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Difference in Mean Change (95% CI); P-value	SMD as Hedges'g (95% CI)	Test for Interaction P-value
	N (a)	(b)	N (a)	(b)	(b)	(b)	(c)
BASELINE HEMOGLOBIN < 8.5 G/DL	51	456.33 (-1.63, 80.23)	23	420.74 (-8.49, 73.93)	6.87 (-18.89, 32.63) 0.5967	0.12 (-0.33, 0.57)	0.6800
>= 8.5 G/DL	35	418.43 (26.09, 105.42)	18	437.61 (9.64, 82.64)	16.45 (-14.17, 47.06) 0.2857	0.28 (-0.25, 0.82)	
BASELINE NTDI-PRO T/W SCORE < 3 POINTS	27	466.22 (-16.73, 67.94)	14	415.07 (0.95, 74.92)	-17.68 (-46.82, 11.45) 0.2267	-0.38 (-1.02, 0.26)	0.0491
>= 3 POINTS	59	429.32 (22.45, 100.37)	27	434.93 (-2.86, 79.08)	25.31 (0.12, 50.50) 0.0489	0.41 (0.00, 0.82)	

(a) Baseline is defined as the last value on or before D1D1.

(b) Estimates are based on an ANCOVA model with treatment group and baseline 6MWT value as continuous covariate.

Hedges g is derived from the estimate of the LS mean difference between Luspatercept and Placebo.

(c) The interaction is p-value is based on an ANCOVA model with treatment group, baseline 6MWT value, the subgroup factor and the treatment-group\*subgroup-factor interaction.

Data after the efficacy cutoff date of a subject will be used to evaluate this endpoint.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-chgsub-ubr700.sas

04AUG2022:03:54:36

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

Protocol: ACE-536-B-THAL-002

Page 5 of 6

Subgroup Analysis of Change from Baseline in 6MWT at Week 48  
ITT Population

Subgroup	Luspatercept (N = 96)		Placebo (N = 49)		Luspatercept vs Placebo		
	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Baseline Mean (SD)	Change from Baseline: LS Mean (95% CI)	Difference in Mean Change (95% CI); P-value	SMD as Hedges'g (95% CI)	Test for Interaction P-value
	N (a)	(b)	N (a)	(b)	(b)	(b)	(c)
BASELINE LIVER IRON							0.3498
< 3 MG/G DRY WEIGHT	32	450.57 ( 83.63) 21.26 (2.85, 39.67)	15	459.89 ( 92.51) -8.72 (-35.63, 18.19)	29.98 (-2.64, 62.61) 0.0707	0.52 (-0.06, 1.11)	
>= 3 TO <= 5 MG/G DRY WEIGHT	20	446.23 ( 89.65) 6.23 (-15.73, 28.20)	10	430.32 ( 77.90) -26.69 (-57.79, 4.41)	32.92 (-5.22, 71.07) 0.0878	0.57 (-0.12, 1.27)	
> 5 TO <= 7 MG/G DRY WEIGHT	12	445.58 ( 59.79) -3.83 (-30.81, 23.15)	1	400.00 (NA) 2.96 (-92.42, 98.33)	-6.79 (-106.23, 92.66) 0.8822	-0.08 (-1.35, 1.18)	
> 7 TO <= 15 MG/G DRY WEIGHT	11	414.94 (164.57) 22.08 (-9.79, 53.94)	8	388.98 ( 56.92) 36.76 (-0.64, 74.15)	-14.68 (-63.94, 34.58) 0.5364	-0.24 (-1.08, 0.60)	
> 15 MG/G DRY WEIGHT	10	427.00 ( 61.92) -19.59 (-53.57, 14.38)	5	407.54 ( 47.21) -8.95 (-57.22, 39.32)	-10.64 (-70.05, 48.76) 0.7031	-0.18 (-1.24, 0.88)	

(a) Baseline is defined as the last value on or before D1D1.

(b) Estimates are based on an ANCOVA model with treatment group and baseline 6MWT value as continuous covariate.

Hedges g is derived from the estimate of the LS mean difference between Luspatercept and Placebo.

(c) The interaction is p-value is based on an ANCOVA model with treatment group, baseline 6MWT value, the subgroup factor and the treatment-group\*subgroup-factor interaction.

Data after the efficacy cutoff date of a subject will be used to evaluate this endpoint.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-chgsub-ebr700.sas

04AUG2022:03:54:36

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

Protocol: ACE-536-B-THAL-002

Page 6 of 6

Subgroup Analysis of Change from Baseline in 6MWT at Week 48  
ITT Population

Subgroup	Luspatercept (N = 96)		Placebo (N = 49)		Luspatercept vs Placebo		
	N (a)	Change from Baseline: Mean (SD) LS Mean (95% CI) (b)	N (a)	Change from Baseline: Mean (SD) LS Mean (95% CI) (b)	Difference in Mean Change (95% CI); P-value (b)	SMD as Hedges'g (95% CI) (b)	Test for Interaction P-value (c)
BASELINE 6MWT							0.6943
<= 450 METERS	49	384.63 (74.46) 24.10 (9.63, 38.57)	28	385.83 (45.31) 15.58 (-3.56, 34.72)	8.52 (-15.47, 32.51) 0.4815	0.16 (-0.29, 0.60)	
> 450 METERS	37	515.43 (53.43) -11.37 (-26.95, 4.21)	13	519.28 (46.24) -29.85 (-56.14, -3.55)	18.48 (-12.09, 49.05) 0.2301	0.33 (-0.22, 0.88)	
BASELINE FACIT-F FS SCORE							0.1076
< 43 POINTS	57	441.22 (104.76) 19.98 (6.01, 33.94)	25	423.85 (73.90) -1.48 (-22.60, 19.64)	21.46 (-3.90, 46.81) 0.0961	0.36 (-0.07, 0.78)	
>= 43 POINTS	29	440.29 (64.07) -6.79 (-24.03, 10.46)	15	423.85 (74.58) 1.91 (-22.12, 25.94)	-8.70 (-38.36, 20.97) 0.5571	-0.17 (-0.77, 0.43)	

(a) Baseline is defined as the last value on or before D1D1.

(b) Estimates are based on an ANCOVA model with treatment group and baseline 6MWT value as continuous covariate.

Hedges g is derived from the estimate of the LS mean difference between Luspatercept and Placebo.

(c) The interaction is p-value is based on an ANCOVA model with treatment group, baseline 6MWT value, the subgroup factor and the treatment-group\*subgroup-factor interaction.

Data after the efficacy cutoff date of a subject will be used to evaluate this endpoint.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ef-chgsub-ibr700.sas

04AUG2022:03:54:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 3

Table 14.1.1.3  
 Summary of Completion Rate of SF-36  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
BASELINE	94/96 (97.9%)	47/49 (95.9%)	141/145 (97.2%)
WEEK 6	78/96 (81.3%)	38/49 (77.6%)	116/145 (80.0%)
WEEK 12	77/96 (80.2%)	45/49 (91.8%)	122/145 (84.1%)
WEEK 18	75/96 (78.1%)	40/49 (81.6%)	115/145 (79.3%)
WEEK 24	78/96 (81.3%)	41/49 (83.7%)	119/145 (82.1%)
WEEK 30	67/96 (69.8%)	35/49 (71.4%)	102/145 (70.3%)
WEEK 36	78/96 (81.3%)	35/49 (71.4%)	113/145 (77.9%)
WEEK 42	68/96 (70.8%)	32/49 (65.3%)	100/145 (69.0%)
WEEK 48	74/96 (77.1%)	41/49 (83.7%)	115/145 (79.3%)
WEEK 54	61/96 (63.5%)	22/49 (44.9%)	83/145 (57.2%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; SF-36 = Medical Outcome Study 36 Item Short Form.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least 50% of the 36 items (i.e., >=18 items) for that visit non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 3

Table 14.1.1.3  
 Summary of Completion Rate of SF-36  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 60	61/96 (63.5%)	26/49 (53.1%)	87/145 (60.0%)
WEEK 66	60/96 (62.5%)	19/49 (38.8%)	79/145 (54.5%)
WEEK 72	54/96 (56.3%)	18/49 (36.7%)	72/145 (49.7%)
WEEK 78	52/96 (54.2%)	15/49 (30.6%)	67/145 (46.2%)
WEEK 84	54/96 (56.3%)	13/49 (26.5%)	67/145 (46.2%)
WEEK 90	50/96 (52.1%)	13/49 (26.5%)	63/145 (43.4%)
WEEK 96	43/96 (44.8%)	8/49 (16.3%)	51/145 (35.2%)
WEEK 102	34/96 (35.4%)	6/49 (12.2%)	40/145 (27.6%)
WEEK 108	30/96 (31.3%)	9/49 (18.4%)	39/145 (26.9%)
WEEK 114	26/96 (27.1%)	9/49 (18.4%)	35/145 (24.1%)

Notes: The ITT population is defined as all randomized subjects.  
 ITT = Intent-to-Treat; SF-36 = Medical Outcome Study 36 Item Short Form.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least 50% of the 36 items (i.e., >=18 items) for that visit non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
Page 3 of 3

Table 14.1.1.3  
Summary of Completion Rate of SF-36  
(ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 120	17/96 (17.7%)	1/49 (2.0%)	18/145 (12.4%)
WEEK 126	6/96 (6.3%)	0/49 (0.0%)	6/145 (4.1%)
WEEK 132	2/96 (2.1%)	0/49 (0.0%)	2/145 (1.4%)

Notes: The ITT population is defined as all randomized subjects.  
ITT = Intent-to-Treat; SF-36 = Medical Outcome Study 36 Item Short Form.  
The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least 50% of the 36 items (i.e., >=18 items) for that visit non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Physical Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	47	
Mean (SD)	47.9 (7.13)	48.2 (7.22)	
Median	48.6	49.1	
Min, Max	28.4, 61.6	32.9, 63.6	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Physical Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	41	
Mean (SD)	49.1 (7.55)	47.3 (6.90)	
Median	49.8	46.5	
Min, Max	27.6, 63.2	34.7, 60.7	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Physical Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	74	41	
Mean (SD)	49.2 (6.51)	48.7 (6.42)	
Median	50.3	47.8	
Min, Max	32.8, 61.0	38.3, 62.5	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Physical Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	92	46	
Week 24	1.3 (-0.03, 2.61)	-0.6 (-2.39, 1.22)	1.9 (-0.29, 4.04) 0.0886 0.3 (-0.05, 0.71)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Physical Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	1.0 (-0.21, 2.11)	0.2 (-1.36, 1.79)	0.7 (-1.14, 2.62) 0.4352 0.2 (-0.24, 0.54)
Overall	1.1 (0.16, 2.12)	-0.2 (-1.56, 1.09)	1.4 (-0.18, 2.92) 0.0829 0.3 (-0.04, 0.67)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Mental Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	94	47	
Mean (SD)	45.9 (9.44)	48.8 (8.34)	
Median	46.5	49.2	
Min, Max	23.3, 60.1	26.5, 63.1	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Mental Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	41	
Mean (SD)	47.9 (9.25)	48.2 (9.12)	
Median	48.9	49.0	
Min, Max	20.9, 61.0	28.8, 61.5	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Mental Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	74	41	
Mean (SD)	47.4 (9.51)	47.4 (8.93)	
Median	47.1	48.2	
Min, Max	27.0, 62.5	25.1, 58.5	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Mental Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	92	46	
Week 24	0.5 (-1.25, 2.28)	-0.7 (-3.18, 1.70)	1.3 (-1.66, 4.17) 0.3970 0.2 (-0.22, 0.54)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 10

Table 14.4.3  
 Observed Scores and Least Square Mean Changes from Baseline of SF-36 by Treatment Group and Visit  
 (ITT Population)

Mental Component Summary

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	0.6 (-1.11, 2.21)	-2.1 (-4.35, 0.24)	2.6 (-0.13, 5.34) 0.0620 0.4 (-0.02, 0.76)
Overall	0.8 (-0.50, 2.16)	-0.8 (-2.66, 1.02)	1.6 (-0.49, 3.78) 0.1298 0.3 (-0.08, 0.63)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	94	27 (28.7)	NE (NE, NE)	47	5 (10.6)	NE (NE, NE)	3.09 (1.18, 8.08)	0.0364	
Region								0.0501	
North America & Europe	59	23 (39.0)	NE (12.4, NE)	28	2 (7.1)	NE (NE, NE)	7.56 (1.75, 32.63)	0.0108	
Middle East	9	1 (11.1)	NE (2.8, NE)	8	1 (12.5)	NE (1.4, NE)	0.02 (<0.01, 26.60)	0.8961	
Asia-Pacific	26	3 (11.5)	NE (NE, NE)	11	2 (18.2)	NE (11.2, NE)	0.60 (0.09, 3.85)	0.6147	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.3098
<=32	35	9 (25.7)	NE (26.6, NE)	11	2 (18.2)	NE (11.2, NE)	1.48 (0.31, 6.97)		0.5098
>32	59	18 (30.5)	NE (NE, NE)	36	3 (8.3)	NE (NE, NE)	3.88 (1.13, 13.27)		0.0341
Splenectomy									0.6468
Yes	33	8 (24.2)	NE (NE, NE)	26	2 (7.7)	NE (NE, NE)	3.73 (0.75, 18.50)		0.1071
No	61	19 (31.1)	NE (26.6, NE)	21	3 (14.3)	NE (NE, NE)	2.24 (0.66, 7.60)		0.2061

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.6919
Male	39	10 (25.6)	NE (NE, NE)	21	1 (4.8)	NE (NE, NE)	4.52 (0.57, 35.60)	0.0988
Female	55	17 (30.9)	NE (26.6, NE)	26	4 (15.4)	NE (NE, NE)	2.79 (0.92, 8.42)	0.1650

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.2225
Beta-Thalassemia	62	22 (35.5)	NE (19.4, NE)	32	3 (9.4)	NE (NE, NE)	4.44 (1.31, 15.09)		0.0201
Hemoglobin	27	3 (11.1)	NE (NE, NE)	11	2 (18.2)	NE (11.2, NE)	0.55 (0.09, 3.53)		0.6147
E/Beta-Thalassemia									
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (2.9, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)		0.2087

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia gene mutation								0.6317
[B0/B0, B+/B+, B+/B0]	67	18 (26.9)	NE (NE, NE)	31	3 (9.7)	NE (NE, NE)	3.79 (1.09, 13.11)	
without Alpha-Thalassemia							0.0592	
[B0/B0, B+/B+, B+/B0]	6	2 (33.3)	NE (2.9, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	
with Alpha-Thalassemia							0.2087	
[B0/B, B+/B] with Alpha gene duplication	21	7 (33.3)	NE (11.9, NE)	12	2 (16.7)	NE (3.0, NE)	1.20 (0.23, 6.40)	0.7951

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept  
 over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess  
 the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	KME (95%CI) (mon) n (%) [a]		Subjects with event N	KME (95%CI) (mon) n (%) [a]		HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Baseline hemoglobin < 8.5 g/dL	54	13 (24.1)	NE (NE, NE)	28	2 (7.1)	NE (NE, NE)	4.91 (1.06, 22.70) 0.0829	0.5247
>= 8.5 g/dL	40	14 (35.0)	NE (19.3, NE)	19	3 (15.8)	NE (NE, NE)	2.03 (0.58, 7.14) 0.2417	
Baseline NTDT-PRO T/W score <3 points	29	2 (6.9)	NE (NE, NE)	14	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.2991	0.9909
>=3 points	65	25 (38.5)	NE (19.3, NE)	33	5 (15.2)	NE (NE, NE)	2.81 (1.06, 7.43) 0.0739	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline liver iron <3 mg/gr dry weight	34	17 (50.0)	19.4 (4.4, NE)	15	1 (6.7)	NE (NE, NE)	12.59 (1.65, 96.12)	0.0117	0.1430
>=3-<=5 mg/gr dry weight	22	5 (22.7)	NE (19.3, NE)	13	1 (7.7)	NE (NE, NE)	3.67 (0.36, 37.25)	0.3751	
>5-<=7 mg/gr dry weight	12	2 (16.7)	NE (10.8, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.6120	
>7-<=15 mg/gr dry weight	14	2 (14.3)	NE (26.6, NE)	9	2 (22.2)	NE (1.4, NE)	0.37 (0.03, 4.14)	0.3502	
>15 mg/gr dry weight	11	1 (9.1)	NE (NE, NE)	5	1 (20.0)	NE (11.2, NE)	0.15 (0.01, 4.17)	0.5922	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline 6MWT <=450	49	12 (24.5)	NE (NE, NE)	31	4 (12.9)	NE (NE, NE)	1.48 (0.46, 4.71)	0.2664	0.0971
>450	45	15 (33.3)	NE (19.3, NE)	16	1 (6.3)	NE (NE, NE)	9.34 (1.20, 72.89)	0.0858	
Baseline FACIT-F FS score <43 points	62	26 (41.9)	NE (12.4, NE)	31	5 (16.1)	NE (NE, NE)	2.79 (1.06, 7.33)	0.0535	0.9913
>=43 points	32	1 (3.1)	NE (NE, NE)	16	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.4576	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b] [c]	[d] [e]	
Overall	94	32 (34.0)	NE (28.1, NE)	47	9 (19.1)	NE (NE, NE)	1.50 (0.71, 3.17) 0.1063		
Region								0.9951	
North America & Europe	59	22 (37.3)	NE (19.9, NE)	28	6 (21.4)	NE (NE, NE)	1.50 (0.59, 3.81) 0.1929		
Middle East	9	2 (22.2)	NE (2.9, NE)	8	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.1549		
Asia-Pacific	26	8 (30.8)	NE (17.1, NE)	11	3 (27.3)	NE (2.8, NE)	1.04 (0.26, 4.12) 0.8662		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.5096
<=32	35	12 (34.3)	28.1 (19.9, NE)	11	2 (18.2)	NE (15.2, NE)	1.95 (0.42, 8.92)		
>32	59	20 (33.9)	NE (17.1, NE)	36	7 (19.4)	NE (NE, NE)	1.27 (0.52, 3.11)		
Splenectomy									0.8543
Yes	33	12 (36.4)	NE (5.7, NE)	26	5 (19.2)	NE (NE, NE)	1.79 (0.62, 5.17)		
No	61	20 (32.8)	NE (28.1, NE)	21	4 (19.0)	NE (15.2, NE)	1.45 (0.49, 4.29)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.2897
Male	39	13 (33.3)	NE (6.9, NE)	21	5 (23.8)	NE (4.4, NE)	0.96 (0.33, 2.83) 0.4167	
Female	55	19 (34.5)	NE (19.9, NE)	26	4 (15.4)	NE (NE, NE)	2.05 (0.69, 6.09) 0.1328	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Beta-thalassemia diagnosis								0.7074	
Beta-Thalassemia	62	22 (35.5)	NE (19.9, NE)	32	4 (12.5)	NE (NE, NE)	1.99 (0.66, 6.01)	0.0280	
Hemoglobin E/Beta-Thalassemia	27	8 (29.6)	NE (17.1, NE)	11	3 (27.3)	NE (2.8, NE)	1.06 (0.27, 4.21)	0.8476	
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (2.9, NE)	4	2 (50.0)	4.4 (4.3, NE)	0.77 (0.10, 5.79)	0.4750	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia gene mutation								0.7243
[B0/B0, B+/B+, B+/B0]	67	23 (34.3)	28.1 (28.1, NE)	31	5 (16.1)	NE (NE, NE)	1.69 (0.63, 4.50)	0.1092
without Alpha-Thalassemia								
[B0/B0, B+/B+, B+/B0]	6	2 (33.3)	NE (2.9, NE)	4	2 (50.0)	4.4 (4.3, NE)	0.77 (0.10, 5.79)	0.4750
with Alpha-Thalassemia								
[B0/B, B+/B] with Alpha gene duplication	21	7 (33.3)	NE (4.4, NE)	12	2 (16.7)	NE (4.2, NE)	1.84 (0.36, 9.48)	0.2602

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	54	21 (38.9)	NE (17.1, NE)	28	6 (21.4)	NE (15.2, NE)	1.47 (0.58, 3.72)	0.8412	
>= 8.5 g/dL	40	11 (27.5)	NE (NE, NE)	19	3 (15.8)	NE (NE, NE)	1.43 (0.39, 5.32)	0.2222	
Baseline NTDT-PRO T/W score <3 points	29	5 (17.2)	NE (NE, NE)	14	2 (14.3)	NE (NE, NE)	0.85 (0.15, 4.83)	0.5924	
>=3 points	65	27 (41.5)	28.1 (8.5, NE)	33	7 (21.2)	NE (15.2, NE)	1.64 (0.71, 3.82)	0.7661	
							0.0964		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron									0.7764
<3 mg/gr dry weight	34	11 (32.4)	NE (19.9, NE)	15	1 (6.7)	NE (NE, NE)	2.53 (0.30, 21.08)		
>=3-<=5 mg/gr dry weight	22	7 (31.8)	NE (4.3, NE)	13	1 (7.7)	NE (NE, NE)	2.95 (0.34, 25.29)		
>5-<=7 mg/gr dry weight	12	5 (41.7)	NE (1.4, NE)	3	2 (66.7)	3.7 (3.0, 4.3)	0.56 (0.06, 4.82)		
>7-<=15 mg/gr dry weight	14	4 (28.6)	NE (2.9, NE)	9	2 (22.2)	NE (2.8, NE)	1.26 (0.22, 7.25)		
>15 mg/gr dry weight	11	5 (45.5)	17.1 (4.3, NE)	5	3 (60.0)	15.2 (1.4, NE)	0.89 (0.20, 3.91)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 16

Table 14.3.3.1.3  
 Summary of Time to First Improvement from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value	Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b][c]	[d][e]
Baseline 6MWT								0.1199
<=450	49	16 (32.7)	NE (19.9, NE)	31	8 (25.8)	NE (15.2, NE)	0.97 (0.41, 2.29)	0.5898
>450	45	16 (35.6)	NE (6.9, NE)	16	1 (6.3)	NE (NE, NE)	5.83 (0.76, 44.81)	0.0417
Baseline FACIT-F FS score								0.3774
<43 points	62	31 (50.0)	19.9 (5.7, NE)	31	8 (25.8)	NE (15.2, NE)	1.66 (0.75, 3.65)	0.0556
>=43 points	32	1 (3.1)	NE (NE, NE)	16	1 (6.3)	NE (NE, NE)	0.58 (0.04, 9.46)	0.6342

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Overall	94	16 (17.0)	NE (NE, NE)	47	10 (21.3)	NE (19.6, NE)	0.68 (0.30, 1.52) 0.2801	
Region								0.7337
North America & Europe	59	7 (11.9)	NE (NE, NE)	28	4 (14.3)	NE (NE, NE)	0.63 (0.17, 2.26) 0.3862	
Middle East	9	3 (33.3)	NE (2.8, NE)	8	4 (50.0)	19.6 (1.4, NE)	2.89 (0.47, 17.54) 0.6872	
Asia-Pacific	26	6 (23.1)	NE (23.0, NE)	11	2 (18.2)	NE (16.8, NE)	1.01 (0.20, 5.09) 0.8723	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.4020
<=32	35	5 (14.3)	NE (NE, NE)	11	3 (27.3)	NE (5.6, NE)	0.19 (0.03, 1.05)		
>32	59	11 (18.6)	NE (NE, NE)	36	7 (19.4)	NE (NE, NE)	0.81 (0.31, 2.11)		
Splenectomy									0.2546
Yes	33	7 (21.2)	NE (NE, NE)	26	5 (19.2)	NE (NE, NE)	1.46 (0.45, 4.71)		
No	61	9 (14.8)	NE (NE, NE)	21	5 (23.8)	NE (16.8, NE)	0.45 (0.15, 1.36)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Sex									0.8347
Male	39	8 (20.5)	NE (NE, NE)	21	5 (23.8)	NE (8.4, NE)	0.80 (0.26, 2.53)		
Female	55	8 (14.5)	NE (NE, NE)	26	5 (19.2)	NE (19.6, NE)	0.71 (0.23, 2.25)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.4872
Beta-Thalassemia	62	7 (11.3)	NE (NE, NE)	32	8 (25.0)	NE (19.6, NE)	0.42 (0.14, 1.26)		0.0206
Hemoglobin E/Beta-Thalassemia	27	7 (25.9)	NE (23.0, NE)	11	2 (18.2)	NE (16.8, NE)	1.29 (0.27, 6.29)		0.6955
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	NE (4.5, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)		0.2457

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Beta-thalassemia gene mutation								0.8319
[B0/B0, B+/B+, B+/B0]	67	12 (17.9)	NE (NE, NE)	31	9 (29.0)	NE (16.8, NE)	0.50 (0.21, 1.21)	
without Alpha-Thalassemia							0.1047	
[B0/B0, B+/B+, B+/B0]	6	2 (33.3)	NE (4.5, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	
with Alpha-Thalassemia							0.2457	
[B0/B, B+/B] with Alpha gene duplication	21	2 (9.5)	NE (NE, NE)	12	1 (8.3)	NE (NE, NE)	1.37 (0.09, 20.52)	0.8794

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept  
 over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess  
 the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	54	11 (20.4)	NE (NE, NE)	28	6 (21.4)	NE (16.8, NE)	0.67 (0.24, 1.86)	0.5154	0.8738
>= 8.5 g/dL	40	5 (12.5)	NE (NE, NE)	19	4 (21.1)	NE (NE, NE)	0.75 (0.19, 3.03)	0.3798	
Baseline NTDT-PRO T/W score <3 points	29	6 (20.7)	NE (NE, NE)	14	5 (35.7)	NE (7.2, NE)	0.57 (0.16, 2.01)	0.2799	0.7951
>=3 points	65	10 (15.4)	NE (NE, NE)	33	5 (15.2)	NE (NE, NE)	0.75 (0.25, 2.23)	0.6947	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline liver iron <3 mg/gr dry weight	34	3 (8.8)	NE (NE, NE)	15	1 (6.7)	NE (NE, NE)	0.85 (0.08, 8.78)	0.3479	
>=3-<=5 mg/gr dry weight	22	4 (18.2)	NE (23.0, NE)	13	3 (23.1)	NE (2.9, NE)	0.44 (0.09, 2.19)	0.8787	
>5-<=7 mg/gr dry weight	12	1 (8.3)	NE (NE, NE)	3	1 (33.3)	NE (7.2, NE)	<0.01 (<0.01, NE)	0.5361	
>7-<=15 mg/gr dry weight	14	3 (21.4)	NE (14.0, NE)	9	2 (22.2)	NE (7.1, NE)	0.46 (0.05, 4.29)	0.0143	
>15 mg/gr dry weight	11	4 (36.4)	NE (7.0, NE)	5	1 (20.0)	NE (16.8, NE)	2.52 (0.22, 28.60)	0.9694	
								0.5273	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Physical Component Summary (Responder Definition=9.4)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	Subjects with event N	n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline 6MWT <=450	49	8 (16.3)	NE (NE, NE)	31	9 (29.0)	NE (16.8, NE)	0.62 (0.23, 1.70)	0.1000	0.4454
>450	45	8 (17.8)	NE (NE, NE)	16	1 (6.3)	NE (NE, NE)	1.43 (0.17, 12.08)	0.4442	
Baseline FACIT-F FS score <43 points	62	9 (14.5)	NE (NE, NE)	31	6 (19.4)	NE (NE, NE)	0.60 (0.21, 1.71)	0.4144	0.5035
>=43 points	32	7 (21.9)	NE (NE, NE)	16	4 (25.0)	NE (16.8, NE)	0.85 (0.23, 3.19)	0.5305	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Overall	94	26 (27.7)	NE (NE, NE)	47	20 (42.6)	NE (9.8, NE)	0.58 (0.32, 1.04) 0.0197	
Region								0.5382
North America & Europe	59	15 (25.4)	NE (NE, NE)	28	11 (39.3)	NE (7.1, NE)	0.47 (0.21, 1.05) 0.0421	
Middle East	9	4 (44.4)	NE (2.9, NE)	8	4 (50.0)	14.5 (1.3, NE)	1.17 (0.28, 4.94) 0.9357	
Asia-Pacific	26	7 (26.9)	NE (25.5, NE)	11	5 (45.5)	NE (2.8, NE)	0.56 (0.17, 1.84) 0.1958	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.3557
<=32	35	11 (31.4)	NE (19.0, NE)	11	5 (45.5)	14.5 (4.2, NE)	0.89 (0.30, 2.64)		
>32	59	15 (25.4)	NE (NE, NE)	36	15 (41.7)	NE (6.4, NE)	0.46 (0.22, 0.95)		
Splenectomy									0.7811
Yes	33	13 (39.4)	NE (9.2, NE)	26	12 (46.2)	14.5 (7.1, NE)	0.74 (0.33, 1.63)		
No	61	13 (21.3)	NE (NE, NE)	21	8 (38.1)	NE (6.4, NE)	0.57 (0.23, 1.44)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Sex									0.5059
Male	39	11 (28.2)	NE (25.5, NE)	21	6 (28.6)	NE (2.8, NE)	0.86 (0.32, 2.36)		0.5597
Female	55	15 (27.3)	NE (NE, NE)	26	14 (53.8)	14.0 (7.1, NE)	0.44 (0.21, 0.92)		0.0152

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.9996
Beta-Thalassemia	62	17 (27.4)	NE (NE, NE)	32	15 (46.9)	11.7 (6.4, NE)	0.51 (0.25, 1.05)		0.0169
Hemoglobin E/Beta-Thalassemia	27	6 (22.2)	NE (25.5, NE)	11	5 (45.5)	NE (2.8, NE)	0.48 (0.14, 1.64)		0.1082
Beta-Thalassemia combined with Alpha-Thalassemia	5	3 (60.0)	12.4 (1.4, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)		0.1269

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b] [c]	[d] [e]	
Beta-thalassemia gene mutation								0.9162	
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	67	17 (25.4)	NE (NE, NE)	31	16 (51.6)	14.0 (5.8, NE)	0.46 (0.23, 0.93)	0.0030	
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	6	3 (50.0)	12.4 (1.4, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.1269	
[B0/B, B+/B] with Alpha gene duplication	21	6 (28.6)	NE (11.5, NE)	12	4 (33.3)	NE (5.8, NE)	0.57 (0.15, 2.14)	0.4628	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	54	16 (29.6)	NE (25.5, NE)	28	11 (39.3)	NE (9.8, NE)	0.64 (0.29, 1.40)	0.1233	0.5366
>= 8.5 g/dL	40	10 (25.0)	NE (NE, NE)	19	9 (47.4)	14.5 (5.7, NE)	0.49 (0.20, 1.22)	0.0747	
Baseline NTDT-PRO T/W score <3 points	29	10 (34.5)	NE (13.3, NE)	14	7 (50.0)	14.5 (6.4, NE)	0.72 (0.27, 1.90)	0.4945	0.5300
>=3 points	65	16 (24.6)	NE (NE, NE)	33	13 (39.4)	NE (5.8, NE)	0.53 (0.25, 1.13)	0.0252	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.6882	
<3 mg/gr dry weight	34	11 (32.4)	NE (19.0, NE)	15	7 (46.7)	11.5 (5.8, NE)	0.64 (0.23, 1.78)	0.1888	
>=3-<=5 mg/gr dry weight	22	6 (27.3)	NE (11.7, NE)	13	8 (61.5)	7.1 (2.7, NE)	0.32 (0.11, 0.94)	0.0209	
>5-<=7 mg/gr dry weight	12	4 (33.3)	NE (4.2, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.4546	
>7-<=15 mg/gr dry weight	14	3 (21.4)	NE (7.0, NE)	9	4 (44.4)	NE (1.4, NE)	0.26 (0.05, 1.44)	0.1950	
>15 mg/gr dry weight	11	1 (9.1)	NE (16.6, NE)	5	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.4292	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable;  
 SF-36 = Medical Outcome Study 36 Item Short Form.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 16

Table 14.3.3.3.3  
 Summary of Time to First Deterioration from Baseline for SF-36 - Subgroup Analysis  
 (ITT Population)

Mental Component Summary (Responder Definition=9.6)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Baseline 6MWT								0.3287
<=450	49	13 (26.5)	NE (NE, NE)	31	12 (38.7)	NE (11.5, NE)	0.70 (0.32, 1.57) 0.1607	
>450	45	13 (28.9)	NE (25.5, NE)	16	8 (50.0)	8.4 (4.2, NE)	0.40 (0.16, 0.99) 0.0323	
Baseline FACIT-F FS score								0.6003
<43 points	62	15 (24.2)	NE (NE, NE)	31	13 (41.9)	NE (5.8, NE)	0.54 (0.25, 1.16) 0.0143	
>=43 points	32	11 (34.4)	NE (11.7, NE)	16	7 (43.8)	NE (6.4, NE)	0.74 (0.28, 1.91) 0.5224	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable; SF-36 = Medical Outcome Study 36 Item Short Form.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 3

Table 14.1.1.2  
 Summary of Completion Rate of FACIT-F  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
BASELINE	96/96 (100.0%)	48/49 (98.0%)	144/145 (99.3%)
WEEK 6	79/96 (82.3%)	38/49 (77.6%)	117/145 (80.7%)
WEEK 12	78/96 (81.3%)	45/49 (91.8%)	123/145 (84.8%)
WEEK 18	75/96 (78.1%)	40/49 (81.6%)	115/145 (79.3%)
WEEK 24	78/96 (81.3%)	42/49 (85.7%)	120/145 (82.8%)
WEEK 30	67/96 (69.8%)	35/49 (71.4%)	102/145 (70.3%)
WEEK 36	78/96 (81.3%)	35/49 (71.4%)	113/145 (77.9%)
WEEK 42	68/96 (70.8%)	32/49 (65.3%)	100/145 (69.0%)
WEEK 48	74/96 (77.1%)	41/49 (83.7%)	115/145 (79.3%)
WEEK 54	61/96 (63.5%)	22/49 (44.9%)	83/145 (57.2%)

Notes: The ITT population is defined as all randomized subjects.  
 FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least 80% of the 40 items (i.e., >=32 items) for that visit non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 3

Table 14.1.1.2  
 Summary of Completion Rate of FACIT-F  
 (ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 60	61/96 (63.5%)	26/49 (53.1%)	87/145 (60.0%)
WEEK 66	60/96 (62.5%)	19/49 (38.8%)	79/145 (54.5%)
WEEK 72	54/96 (56.3%)	18/49 (36.7%)	72/145 (49.7%)
WEEK 78	52/96 (54.2%)	15/49 (30.6%)	67/145 (46.2%)
WEEK 84	54/96 (56.3%)	13/49 (26.5%)	67/145 (46.2%)
WEEK 90	50/96 (52.1%)	13/49 (26.5%)	63/145 (43.4%)
WEEK 96	43/96 (44.8%)	8/49 (16.3%)	51/145 (35.2%)
WEEK 102	34/96 (35.4%)	6/49 (12.2%)	40/145 (27.6%)
WEEK 108	30/96 (31.3%)	9/49 (18.4%)	39/145 (26.9%)
WEEK 114	26/96 (27.1%)	9/49 (18.4%)	35/145 (24.1%)

Notes: The ITT population is defined as all randomized subjects.  
 FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat.  
 The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least 80% of the 40 items (i.e., >=32 items) for that visit non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
Page 3 of 3

Table 14.1.1.2  
Summary of Completion Rate of FACIT-F  
(ITT Population)

Timepoint	Luspatercept	Placebo	Overall
	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)	Number of subjects/ Number of subjects in the ITT (%)
WEEK 120	17/96 (17.7%)	1/49 (2.0%)	18/145 (12.4%)
WEEK 126	6/96 (6.3%)	0/49 (0.0%)	6/145 (4.1%)
WEEK 132	2/96 (2.1%)	0/49 (0.0%)	2/145 (1.4%)

Notes: The ITT population is defined as all randomized subjects.  
FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat.  
The completion rate is calculated as [Number of subjects with an evaluable assessment at the timepoint] / [Number of subjects in the ITT population]. The evaluable assessment is defined as at least 80% of the 40 items (i.e., >=32 items) for that visit non-missing.

Program Source: ...06-Programing\A-Primary\Programs\Table\t\_completion.sas

13MAY2022 02:36

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Total Score

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	96	48	
Mean (SD)	114.6 (23.85)	115.0 (20.65)	
Median	119.0	117.0	
Min, Max	62.0, 155.8	72.0, 151.7	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Total Score

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	42	
Mean (SD)	119.4 (22.97)	114.9 (20.85)	
Median	122.5	120.2	
Min, Max	53.0, 159.0	74.0, 147.5	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Total Score

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	74	41	
Mean (SD)	119.5 (22.22)	116.8 (18.78)	
Median	121.8	120.7	
Min, Max	57.0, 159.0	81.0, 149.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Total Score

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	2.0 (-1.38, 5.43)	-2.7 (-7.32, 1.95)	4.7 (-0.79, 10.23) 0.0927 0.3 (-0.06, 0.70)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Total Score

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	2.4 (-0.89, 5.76)	-1.2 (-5.68, 3.31)	3.6 (-1.73, 8.97) 0.1828 0.3 (-0.13, 0.64)
Overall	2.2 (-0.70, 5.01)	-1.9 (-5.77, 1.99)	4.0 (-0.48, 8.57) 0.0793 0.3 (-0.04, 0.66)

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Physical Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Baseline			
N [a]	96	48	
Mean (SD)	22.8 (4.00)	23.0 (3.81)	
Median	24.0	24.0	
Min, Max	11.0, 28.0	13.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Physical Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Week 24			
N [a]	78	42	
Mean (SD)	23.0 (4.56)	22.5 (3.76)	
Median	24.0	22.5	
Min, Max	10.0, 28.0	13.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Physical Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Week 48			
N [a]	74	41	
Mean (SD)	23.6 (4.26)	23.2 (3.49)	
Median	25.0	23.0	
Min, Max	10.0, 28.0	15.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 9 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Physical Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	-0.1 (-0.87, 0.74)	-0.4 (-1.51, 0.69)	0.4 (-0.96, 1.66) 0.5967 0.1 (-0.28, 0.48)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 10 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Physical Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	0.3 (-0.49, 1.14)	-0.3 (-1.42, 0.80)	0.6 (-0.68, 1.96) 0.3400 0.2 (-0.20, 0.57)
Overall	0.0 (-0.65, 0.68)	-0.3 (-1.19, 0.62)	0.3 (-0.76, 1.35) 0.5820 0.1 (-0.25, 0.45)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 11 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Social/Family Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	96	48	
Mean (SD)	19.5 (5.70)	19.3 (4.44)	
Median	20.0	19.0	
Min, Max	4.7, 28.0	10.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 12 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Social/Family Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Week 24			
N [a]	78	42	
Mean (SD)	19.7 (5.38)	19.9 (5.16)	
Median	20.0	21.0	
Min, Max	2.3, 28.0	8.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 13 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Social/Family Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	74	41	
Mean (SD)	18.8 (5.69)	19.4 (4.46)	
Median	19.5	20.0	
Min, Max	6.0, 28.0	10.0, 26.8	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 14 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Social/Family Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	0.0 (-0.95, 0.88)	-0.2 (-1.43, 1.07)	0.1 (-1.36, 1.66) 0.8456 0.0 (-0.34, 0.42)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 15 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Social/Family Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	-0.5 (-1.26, 0.30)	0.1 (-1.00, 1.11)	-0.5 (-1.80, 0.73) 0.4027 -0.2 (-0.55, 0.22)
Overall	-0.2 (-0.81, 0.51)	-0.4 (-1.27, 0.52)	0.2 (-0.83, 1.28) 0.6715 0.1 (-0.28, 0.43)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 16 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Emotional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	96	48	
Mean (SD)	18.2 (3.69)	18.2 (3.25)	
Median	19.0	18.0	
Min, Max	8.0, 24.0	10.0, 23.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 17 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Emotional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	42	
Mean (SD)	19.0 (3.52)	18.1 (3.53)	
Median	19.0	18.0	
Min, Max	9.0, 24.0	10.0, 24.0	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 18 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Emotional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	74	41	
Mean (SD)	19.1 (3.66)	18.4 (3.71)	
Median	20.0	19.0	
Min, Max	8.0, 24.0	12.0, 24.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 19 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Emotional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	0.3 (-0.26, 0.92)	-0.3 (-1.07, 0.54)	0.6 (-0.37, 1.56) 0.2222 0.2 (-0.15, 0.61)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 20 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Emotional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	0.4 (-0.22, 0.97)	-0.1 (-0.86, 0.75)	0.4 (-0.53, 1.40) 0.3731 0.2 (-0.21, 0.56)
Overall	0.4 (-0.07, 0.90)	-0.1 (-0.72, 0.59)	0.5 (-0.29, 1.26) 0.2156 0.2 (-0.13, 0.57)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 21 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Functional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	96	48	
Mean (SD)	17.8 (5.60)	18.2 (5.09)	
Median	18.0	18.0	
Min, Max	3.0, 28.0	7.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 22 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Functional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	42	
Mean (SD)	18.6 (5.18)	17.4 (5.81)	
Median	18.0	19.0	
Min, Max	7.0, 28.0	6.0, 28.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 23 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Functional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Week 48			
N [a]	74	41	
Mean (SD)	18.2 (5.00)	17.8 (4.94)	
Median	18.0	18.0	
Min, Max	4.0, 28.0	8.0, 26.0	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 24 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Functional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	0.4 (-0.41, 1.16)	-1.4 (-2.47, -0.32)	1.8 (0.48, 3.07) 0.0075 0.5 (0.13, 0.90)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 25 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Functional Well-Being

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	0.0 (-0.75, 0.74)	-0.6 (-1.66, 0.36)	0.6 (-0.57, 1.86) 0.2959 0.2 (-0.18, 0.59)
Overall	0.2 (-0.45, 0.79)	-0.9 (-1.79, -0.11)	1.1 (0.13, 2.11) 0.0273 0.4 (0.04, 0.75)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 26 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

FACT-G

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	96	48	
Mean (SD)	78.3 (15.45)	78.7 (12.63)	
Median	80.0	78.5	
Min, Max	42.0, 105.8	51.0, 102.7	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 27 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

FACT-G

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	42	
Mean (SD)	80.3 (15.31)	78.0 (14.03)	
Median	81.0	81.0	
Min, Max	44.0, 108.0	46.0, 103.0	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 28 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

FACT-G

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48			
N [a]	74	41	
Mean (SD)	79.6 (15.03)	78.8 (12.41)	
Median	81.0	79.7	
Min, Max	39.0, 108.0	57.0, 101.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 29 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

FACT-G

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	0.5 (-1.73, 2.72)	-2.3 (-5.38, 0.68)	2.8 (-0.77, 6.46) 0.1222 0.3 (-0.08, 0.68)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 30 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

FACT-G

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	0.2 (-1.90, 2.26)	-1.1 (-3.88, 1.74)	1.2 (-2.10, 4.59) 0.4615 0.1 (-0.24, 0.53)
Overall	0.3 (-1.57, 2.15)	-1.9 (-4.38, 0.67)	2.1 (-0.81, 5.11) 0.1535 0.3 (-0.10, 0.60)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 31 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Fatigue-related Symptoms

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Baseline			
N [a]	96	48	
Mean (SD)	36.4 (10.04)	36.3 (9.50)	
Median	39.0	38.0	
Min, Max	1.0, 51.0	16.0, 50.0	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 32 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Fatigue-related Symptoms

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 24			
N [a]	78	42	
Mean (SD)	39.1 (9.21)	36.9 (8.88)	
Median	42.0	39.5	
Min, Max	9.0, 51.0	11.0, 50.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 33 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Fatigue-related Symptoms

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
Week 48			
N [a]	74	41	
Mean (SD)	39.9 (8.44)	38.0 (7.67)	
Median	42.0	41.0	
Min, Max	12.0, 52.0	22.0, 50.0	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 34 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Fatigue-related Symptoms

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	
LS mean change from baseline (95% CI) [b]			
N [c]	94	47	
Week 24	1.8 (0.24, 3.40)	-0.2 (-2.32, 2.00)	2.0 (-0.60, 4.55) 0.1314 0.3 (-0.09, 0.67)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.  
 [b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.  
 [c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmr.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 35 of 35

Table 14.4.2  
 Observed Scores and Least Square Mean Changes from Baseline of FACIT-F by Treatment Group and Visit  
 (ITT Population)

Fatigue-related Symptoms

	Luspatercept (n = 96)	Placebo (n = 49)	Luspatercept vs. Placebo
	Descriptive statistics/ MMRM estimates	Descriptive statistics/ MMRM estimates	LS Mean Difference (95% CI) p-value Hedges' g (95% CI)
Week 48	2.4 (0.86, 3.91)	0.2 (-1.86, 2.28)	2.2 (-0.29, 4.65) 0.0826 0.3 (-0.05, 0.72)
Overall	2.1 (0.82, 3.33)	0.1 (-1.56, 1.84)	1.9 (-0.04, 3.92) 0.0545 0.3 (-0.01, 0.69)

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; ITT = Intent-to-Treat; LS = Least Square; Max = Maximum; Min = Minimum; MMRM = Mixed Model for Repeated Measurements; SD = Standard Deviation.

[a] Number of subjects in the analysis at the specific assessment timepoint.

[b] Estimates based on a MMRM with change from baseline as dependent variable, treating treatment, visit week, stratification factors, and treatment\*visit week interaction as fixed effects, baseline PRO score as covariate, and visit week as repeated measure. An unstructured covariance matrix was used in the model. Only the timepoints with at least 10 subjects with non-missing score in each treatment group are included.

[c] Number of subjects included in the model.

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_chg\_mmrn.sas

18MAY2022 13:38

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Overall	96	28 (29.2)	NE (NE, NE)	48	6 (12.5)	NE (NE, NE)	2.53	(1.04, 6.15)	0.0431
Region									0.3385
North America & Europe	60	21 (35.0)	NE (11.5, NE)	29	4 (13.8)	NE (13.3, NE)	3.52	(1.20, 10.37)	0.0590
Middle East	9	2 (22.2)	NE (2.8, NE)	8	0 (0.0)	NE (NE, NE)	>9999	(<0.01, NE)	0.1864
Asia-Pacific	27	5 (18.5)	NE (NE, NE)	11	2 (18.2)	NE (15.2, NE)	0.95	(0.18, 5.20)	0.9326

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.2084
<=32	35	9 (25.7)	NE (NE, NE)	12	2 (16.7)	NE (15.2, NE)	1.06 (0.22, 5.20)		
>32	61	19 (31.1)	NE (NE, NE)	36	4 (11.1)	NE (NE, NE)	3.50 (1.18, 10.35)		
Splenectomy									0.2619
Yes	34	8 (23.5)	NE (NE, NE)	26	4 (15.4)	NE (NE, NE)	1.39 (0.41, 4.75)		
No	62	20 (32.3)	NE (NE, NE)	22	2 (9.1)	NE (15.2, NE)	4.69 (1.08, 20.30)		

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.9092
Male	40	10 (25.0)	NE (NE, NE)	22	2 (9.1)	NE (NE, NE)	2.54 (0.54, 11.86) 0.1258	
Female	56	18 (32.1)	NE (19.9, NE)	26	4 (15.4)	NE (NE, NE)	2.75 (0.92, 8.22) 0.1336	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.2659
Beta-Thalassemia	63	20 (31.7)	NE (NE, NE)	33	3 (9.1)	NE (NE, NE)	4.06 (1.19, 13.80)		0.0196
Hemoglobin	28	5 (17.9)	NE (NE, NE)	11	2 (18.2)	NE (15.2, NE)	0.95 (0.17, 5.19)		0.9185
E/Beta-Thalassemia									
Beta-Thalassemia combined with Alpha-Thalassemia	5	3 (60.0)	18.3 (3.2, NE)	4	1 (25.0)	NE (4.3, NE)	1.76 (0.18, 16.98)		0.6247

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo			
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value	
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d]	[e]
Beta-thalassemia gene mutation										0.8817
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	69	20 (29.0)	NE (NE, NE)	32	4 (12.5)	NE (NE, NE)	2.40	(0.81, 7.06)		0.0901
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	6	3 (50.0)	18.3 (3.2, NE)	4	1 (25.0)	NE (4.3, NE)	1.76	(0.18, 16.98)		0.6247
[B0/B, B+/B] with Alpha gene duplication	21	5 (23.8)	NE (9.9, NE)	12	1 (8.3)	NE (13.3, NE)	4.07	(0.45, 36.74)		0.3239

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Baseline hemoglobin < 8.5 g/dL	55	17 (30.9)	NE (NE, NE)	28	4 (14.3)	NE (NE, NE)	1.44 (0.46, 4.49)	0.1753	0.2289
>= 8.5 g/dL	41	11 (26.8)	NE (NE, NE)	20	2 (10.0)	NE (13.3, NE)	5.44 (1.15, 25.86)	0.1289	
Baseline NTDT-PRO T/W score <3 points	30	2 (6.7)	NE (NE, NE)	14	1 (7.1)	NE (NE, NE)	0.32 (0.02, 6.46)	0.8329	0.3065
>=3 points	66	26 (39.4)	NE (11.2, NE)	34	5 (14.7)	NE (15.2, NE)	2.98 (1.14, 7.82)	0.0238	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value		Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b]	[c]	[d] [e]
Baseline liver iron									0.1520
<3 mg/gr dry weight	35	13 (37.1)	NE (9.9, NE)	16	1 (6.3)	NE (13.3, NE)	8.50	(1.10, 65.87)	
>=3-<=5 mg/gr dry weight	23	5 (21.7)	NE (NE, NE)	13	1 (7.7)	NE (NE, NE)	2.48	(0.27, 22.60)	0.0320
>5-<=7 mg/gr dry weight	12	5 (41.7)	NE (1.4, NE)	3	1 (33.3)	NE (4.3, NE)	0.15	(0.01, 2.66)	0.2513
>7-<=15 mg/gr dry weight	14	4 (28.6)	NE (7.2, NE)	9	1 (11.1)	NE (2.8, NE)	1.89	(0.19, 19.18)	0.4948
>15 mg/gr dry weight	11	1 (9.1)	NE (NE, NE)	5	2 (40.0)	NE (4.3, NE)	<0.01	(<0.01, NE)	0.3564
									0.2087

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 8

Table 14.3.3.1.2  
 Summary of Time to First Improvement from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Baseline 6MWT								
<=450	51	14 (27.5)	NE (NE, NE)	31	6 (19.4)	NE (NE, NE)	1.34 (0.51, 3.51) 0.4618	0.9892
>450	45	14 (31.1)	NE (18.3, NE)	17	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.0195	
Baseline FACIT-F FS score								
<43 points	64	26 (40.6)	NE (11.2, NE)	32	6 (18.8)	NE (15.2, NE)	2.32 (0.95, 5.68) 0.0654	0.9918
>=43 points	32	2 (6.3)	NE (NE, NE)	16	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE) 0.3425	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 1 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Overall	96	21 (21.9)	NE (NE, NE)	48	12 (25.0)	NE (NE, NE)	0.76 (0.37, 1.56)	0.4434	
Region								0.3572	
North America & Europe	60	9 (15.0)	NE (NE, NE)	29	6 (20.7)	NE (NE, NE)	0.44 (0.15, 1.32)	0.2771	
Middle East	9	4 (44.4)	22.5 (1.5, NE)	8	4 (50.0)	14.5 (2.8, NE)	1.20 (0.26, 5.42)	0.8686	
Asia-Pacific	27	8 (29.6)	NE (21.0, NE)	11	2 (18.2)	NE (5.6, NE)	1.65 (0.35, 7.76)	0.5226	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 2 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Age									0.5379
<=32	35	6 (17.1)	NE (NE, NE)	12	4 (33.3)	NE (7.0, NE)	0.56 (0.15, 2.00)		
>32	61	15 (24.6)	NE (NE, NE)	36	8 (22.2)	NE (NE, NE)	0.91 (0.38, 2.16)		
Splenectomy									0.7241
Yes	34	10 (29.4)	NE (22.5, NE)	26	7 (26.9)	NE (14.5, NE)	0.99 (0.37, 2.64)		
No	62	11 (17.7)	NE (NE, NE)	22	5 (22.7)	NE (11.5, NE)	0.75 (0.26, 2.18)		

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 3 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Sex								0.7718
Male	40	12 (30.0)	NE (NE, NE)	22	6 (27.3)	NE (7.1, NE)	0.92 (0.34, 2.48)	0.8556
Female	56	9 (16.1)	NE (NE, NE)	26	6 (23.1)	NE (14.5, NE)	0.64 (0.23, 1.82)	0.3816

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 4 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		Test for interaction p-value [d] [e]
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]		
Beta-thalassemia diagnosis									0.2631
Beta-Thalassemia	63	10 (15.9)	NE (NE, NE)	33	10 (30.3)	NE (11.5, NE)	0.44 (0.18, 1.06)		0.0501
Hemoglobin E/Beta-Thalassemia	28	9 (32.1)	NE (19.6, NE)	11	2 (18.2)	NE (5.6, NE)	1.87 (0.40, 8.65)		0.4154
Beta-Thalassemia combined with Alpha-Thalassemia	5	2 (40.0)	22.5 (1.4, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)		0.2945

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 5 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	Subjects with event		KME (95%CI) (mon)	Subjects with event		KME (95%CI) (mon)	HR (95% CI) p-value	Test for interaction p-value
	N	n (%)	[a]	N	n (%)	[a]	[b] [c]	[d] [e]
Beta-thalassemia gene mutation								0.9857
[B0/B0, B+/B+, B+/B0] without Alpha-Thalassemia	69	16 (23.2)	NE (NE, NE)	32	10 (31.3)	NE (14.5, NE)	0.69 (0.31, 1.53)	0.2965
[B0/B0, B+/B+, B+/B0] with Alpha-Thalassemia	6	2 (33.3)	22.5 (1.4, NE)	4	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.2945
[B0/B, B+/B] with Alpha gene duplication	21	3 (14.3)	NE (NE, NE)	12	2 (16.7)	NE (5.6, NE)	0.65 (0.10, 4.11)	0.8583

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 6 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]
Baseline hemoglobin < 8.5 g/dL	55	12 (21.8)	NE (NE, NE)	28	7 (25.0)	NE (NE, NE)	0.71 (0.28, 1.83) 0.4894	0.8807
>= 8.5 g/dL	41	9 (22.0)	NE (24.7, NE)	20	5 (25.0)	NE (14.5, NE)	0.84 (0.28, 2.52) 0.7305	
Baseline NTDT-PRO T/W score <3 points	30	9 (30.0)	NE (21.0, NE)	14	5 (35.7)	NE (7.0, NE)	0.90 (0.30, 2.71) 0.7934	0.8413
>=3 points	66	12 (18.2)	NE (NE, NE)	34	7 (20.6)	NE (NE, NE)	0.76 (0.30, 1.95) 0.5151	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 7 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo		
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b] [c]	Test for interaction p-value [d] [e]	
Baseline liver iron								0.9003	
<3 mg/gr dry weight	35	4 (11.4)	NE (NE, NE)	16	2 (12.5)	NE (11.5, NE)	0.91 (0.17, 5.00)	0.8885	
>=3-<=5 mg/gr dry weight	23	6 (26.1)	NE (19.6, NE)	13	7 (53.8)	7.1 (4.2, NE)	0.22 (0.06, 0.78)	0.0422	
>5-<=7 mg/gr dry weight	12	3 (25.0)	NE (7.8, NE)	3	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.5501	
>7-<=15 mg/gr dry weight	14	2 (14.3)	NE (22.5, NE)	9	2 (22.2)	NE (2.8, NE)	0.55 (0.07, 4.25)	0.6081	
>15 mg/gr dry weight	11	5 (45.5)	24.7 (4.4, NE)	5	0 (0.0)	NE (NE, NE)	>9999 (<0.01, NE)	0.1025	

Notes: The ITT population is defined as all randomized subjects.  
 CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.  
 [a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).  
 [b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.  
 [c] Unstratified Log-rank test.  
 [d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.  
 [e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).  
 Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

Celgene Corporation  
 Protocol: ACE-536-B-THAL-002

Database Cutoff Date: 14SEP2020  
 Page 8 of 8

Table 14.3.3.3.2  
 Summary of Time to First Deterioration from Baseline for FACIT-F - Subgroup Analysis  
 (ITT Population)

Total Score (Responder Definition=24)

Subgroup	Luspatercept (n = 96)			Placebo (n = 49)			Luspatercept vs. Placebo	
	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	N	Subjects with event n (%)	KME (95%CI) (mon) [a]	HR (95% CI) p-value [b][c]	Test for interaction p-value [d][e]
Baseline 6MWT								
<=450	51	14 (27.5)	NE (NE, NE)	31	8 (25.8)	NE (NE, NE)	1.05 (0.44, 2.52) 0.9566	0.2808
>450	45	7 (15.6)	NE (NE, NE)	17	4 (23.5)	NE (7.1, NE)	0.54 (0.16, 1.90) 0.3381	
Baseline FACIT-F FS score								
<43 points	64	14 (21.9)	NE (NE, NE)	32	6 (18.8)	NE (NE, NE)	1.03 (0.39, 2.71) 0.9414	0.2705
>=43 points	32	7 (21.9)	NE (NE, NE)	16	6 (37.5)	NE (7.0, NE)	0.50 (0.17, 1.54) 0.2489	

Notes: The ITT population is defined as all randomized subjects.

CI = Confidence Interval; FACIT-F = Functional Assessment of Chronic Illness Therapy-Fatigue; HR = hazard ratio; ITT = Intent-to-Treat; KME = Kaplan-Meier Estimate; NE = Not evaluable.

[a] KME of median time. Two-sided 95% CI is computed by Brookmeyer and Crowley method (log-log transformation).

[b] Unstratified Cox proportional hazard model, with baseline PRO score as a covariate. Hazard Ratio is Luspatercept over Placebo.

[c] Unstratified Log-rank test.

[d] Unstratified Cox proportional hazard model with treatment, subgroup, and treatment\*subgroup interaction to assess the significance of the interaction between treatment and the subgroup, with baseline PRO score as a covariate.

[e] P-values <0.05 are indicated by 1 asterisk (indicates potential effect modification).

Program Source: ...06-Programing\B-Secondary\Programs\Table\t\_tte\_subgrp.sas

27MAY2022 07:27

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

Protocol: ACE-536-B-THAL-002

Page 1 of 6

Time to first Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
TOTAL SUBJECTS WITH AN EVENT	96	96 (100.0)	0.10 (0.07, 0.13)	49	48 ( 98.0)	0.76 (0.46, 0.89)	1.47 (1.03, 2.12)	0.0433
Musculoskeletal and connective tissue disorders	96	76 ( 79.2)	0.72 (0.13, 2.07)	49	28 ( 57.1)	8.51 (1.74, N.A.)	1.85 (1.19, 2.86)	0.0065
Bone pain	96	35 ( 36.5)	N.A.	49	3 ( 6.1)	N.A.	7.11 (2.18, 23.15)	0.0001
Arthralgia	96	28 ( 29.2)	N.A.	49	7 ( 14.3)	N.A.	1.96 (0.85, 4.50)	0.1064
Back pain	96	27 ( 28.1)	N.A. (28.91, N.A.)	49	5 ( 10.2)	N.A.	2.46 (0.94, 6.43)	0.0580
Pain in extremity	96	16 ( 16.7)	N.A.	49	5 ( 10.2)	N.A.	1.37 (0.50, 3.77)	0.5422
Myalgia	96	11 ( 11.5)	N.A.	49	5 ( 10.2)	N.A.	0.99 (0.34, 2.87)	0.9912
Infections and infestations	96	71 ( 74.0)	7.62 (5.52, 10.87)	49	33 ( 67.3)	7.36 (3.25, 10.32)	0.96 (0.63, 1.46)	0.8469

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocae-ebr700.sas

04AUG2022:03:50:51

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

Protocol: ACE-536-B-THAL-002

Page 2 of 6

Time to first Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
Upper respiratory tract infection	96	20 ( 20.8)	N.A.	49	11 ( 22.4)	N.A.	0.83 (0.39, 1.74)	0.6161
Pharyngitis	96	19 ( 19.8)	N.A.	49	7 ( 14.3)	N.A.	1.03 (0.42, 2.49)	0.9544
Influenza	96	12 ( 12.5)	N.A.	49	5 ( 10.2)	N.A.	1.12 (0.39, 3.21)	0.8295
Gastroenteritis	96	10 ( 10.4)	N.A.	49	3 ( 6.1)	N.A.	1.30 (0.35, 4.84)	0.6979
Rhinitis	96	8 ( 8.3)	N.A.	49	5 ( 10.2)	N.A.	0.65 (0.21, 2.00)	0.4477
Tonsillitis	96	2 ( 2.1)	N.A.	49	6 ( 12.2)	N.A.	0.13 (0.03, 0.64)	0.0033
General disorders and administration site conditions	96	53 ( 55.2)	14.52 (7.79, N.A.)	49	22 ( 44.9)	19.38 (4.70, N.A.)	1.18 (0.72, 1.95)	0.5075
Fatigue	96	16 ( 16.7)	N.A.	49	10 ( 20.4)	N.A.	0.70 (0.31, 1.57)	0.3878

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-T/PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocae-ebr700.sas

04AUG2022:03:50:51

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

Protocol: ACE-536-B-THAL-002

Page 3 of 6

Time to first Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
Influenza like illness	96	16 ( 16.7)	N.A.	49	3 ( 6.1)	N.A.	2.23 (0.65, 7.72)	0.1916
Pyrexia	96	14 ( 14.6)	N.A.	49	9 ( 18.4)	N.A.	0.63 (0.27, 1.46)	0.2754
Asthenia	96	13 ( 13.5)	N.A.	49	5 ( 10.2)	N.A.	1.02 (0.36, 2.91)	0.9736
Gastrointestinal disorders	96	47 ( 49.0)	21.91 (10.91, N.A.)	49	22 ( 44.9)	N.A. (5.52, N.A.)	0.88 (0.53, 1.47)	0.6307
Diarrhoea	96	16 ( 16.7)	N.A.	49	6 ( 12.2)	N.A.	1.12 (0.44, 2.89)	0.8098
Toothache	96	12 ( 12.5)	N.A.	49	1 ( 2.0)	N.A.	5.21 (0.67, 40.22)	0.0773
Nausea	96	10 ( 10.4)	N.A.	49	6 ( 12.2)	N.A.	0.79 (0.29, 2.17)	0.6432
Abdominal pain	96	9 ( 9.4)	N.A.	49	5 ( 10.2)	N.A.	0.53 (0.17, 1.68)	0.2742

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-T/PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocae-ebr700.sas

04AUG2022:03:50:51

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

Protocol: ACE-536-B-THAL-002

Page 4 of 6

Time to first Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
Nervous system disorders	96	44 ( 45.8)	28.55 (9.43, N.A.)	49	16 ( 32.7)	N.A.	1.42 (0.79, 2.53)	0.2385
Headache	96	29 ( 30.2)	N.A.	49	10 ( 20.4)	N.A.	1.49 (0.73, 3.07)	0.2688
Respiratory, thoracic and mediastinal disorders	96	40 ( 41.7)	N.A. (16.76, N.A.)	49	15 ( 30.6)	N.A.	1.25 (0.69, 2.27)	0.4718
Oropharyngeal pain	96	19 ( 19.8)	N.A.	49	6 ( 12.2)	N.A.	1.47 (0.59, 3.71)	0.4065
Cough	96	17 ( 17.7)	N.A.	49	1 ( 2.0)	N.A.	7.24 (0.96, 54.68)	0.0248
Vascular disorders	96	37 ( 38.5)	N.A. (25.07, N.A.)	49	10 ( 20.4)	N.A.	1.97 (0.98, 3.97)	0.0530
Prehypertension	96	22 ( 22.9)	N.A.	49	7 ( 14.3)	N.A.	1.60 (0.68, 3.75)	0.2762
Hypertension	96	19 ( 19.8)	N.A.	49	1 ( 2.0)	N.A.	9.32 (1.25, 69.66)	0.0079

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-T/PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocae-eb700.sas

04AUG2022:03:50:51

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

Protocol: ACE-536-B-THAL-002

Page 5 of 6

Time to first Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
Injury, poisoning and procedural complications	96	26 ( 27.1)	N.A.	49	7 ( 14.3)	N.A.	1.68 (0.72, 3.88)	0.2220
Skin and subcutaneous tissue disorders	96	26 ( 27.1)	N.A.	49	5 ( 10.2)	N.A.	2.59 (0.99, 6.77)	0.0434
Reproductive system and breast disorders	96	24 ( 25.0)	N.A.	49	6 ( 12.2)	N.A.	1.60 (0.65, 3.96)	0.3028
Menstruation irregular	96	12 ( 12.5)	N.A.	49	3 ( 6.1)	N.A.	1.52 (0.42, 5.47)	0.5176
Metabolism and nutrition disorders	96	23 ( 24.0)	N.A.	49	13 ( 26.5)	N.A.	0.70 (0.35, 1.40)	0.3157
Iron overload	96	10 ( 10.4)	N.A.	49	5 ( 10.2)	N.A.	0.91 (0.31, 2.66)	0.8556
Psychiatric disorders	96	17 ( 17.7)	N.A.	49	2 ( 4.1)	N.A.	3.69 (0.85, 16.07)	0.0628
Insomnia	96	11 ( 11.5)	N.A.	49	1 ( 2.0)	N.A.	4.92 (0.63, 38.29)	0.0920

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-T/PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocae-ebr700.sas

04AUG2022:03:50:51

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

Protocol: ACE-536-B-THAL-002

Page 6 of 6

Time to first Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
Blood and lymphatic system disorders	96	14 ( 14.6)	N.A.	49	7 ( 14.3)	N.A. (21.42, N.A.)	0.86 (0.34, 2.15)	0.7440
Eye disorders	96	13 ( 13.5)	N.A.	49	4 ( 8.2)	N.A.	1.44 (0.47, 4.46)	0.5220
Investigations	96	13 ( 13.5)	N.A.	49	4 ( 8.2)	N.A.	1.42 (0.46, 4.39)	0.5393
Ear and labyrinth disorders	96	12 ( 12.5)	N.A.	49	7 ( 14.3)	N.A.	0.75 (0.29, 1.93)	0.5568
Cardiac disorders	96	11 ( 11.5)	N.A.	49	9 ( 18.4)	N.A.	0.42 (0.17, 1.03)	0.0502
Palpitations	96	6 ( 6.3)	N.A.	49	6 ( 12.2)	N.A.	0.36 (0.11, 1.14)	0.0705

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocae-ebr700.sas

04AUG2022:03:50:51

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to first Serious Adverse Events by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
TOTAL SUBJECTS WITH AN EVENT	96	11 ( 11.5)	N.A.	49	12 ( 24.5)	N.A. (18.00, N.A.)	0.29 (0.12, 0.69)	0.0032
Injury, poisoning and procedural complications	96	5 ( 5.2)	N.A.	49	1 ( 2.0)	N.A.	1.76 (0.20, 15.77)	0.6102
Traumatic fracture	96	5 ( 5.2)	N.A.	49	1 ( 2.0)	N.A.	1.76 (0.20, 15.77)	0.6102
Infections and infestations	96	2 ( 2.1)	N.A.	49	6 ( 12.2)	N.A. (19.48, N.A.)	0.11 (0.02, 0.57)	0.0017

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsa-ebr700.sas

04AUG2022:03:51:22

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to first Adverse Events CTC Grade >=3 by SOC/PT  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
TOTAL SUBJECTS WITH AN EVENT	96	27 ( 28.1)	N.A.	49	12 ( 24.5)	N.A. (16.62, N.A.)	1.07 (0.54, 2.14)	0.8419
Musculoskeletal and connective tissue disorders	96	7 ( 7.3)	N.A.	49	0	N.E.	N.E.	0.0562
Metabolism and nutrition disorders	96	5 ( 5.2)	N.A.	49	0	N.E.	N.E.	0.1217
Infections and infestations	96	1 ( 1.0)	N.A.	49	6 ( 12.2)	N.A. (19.48, N.A.)	0.05 (0.01, 0.42)	0.0002

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/tables/rt-ae-ttfsocsa-ebr700.sas

04AUG2022:03:51:16

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Malignancy  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY MALIGNANCY	96	0	N.E.	49	2 ( 4.1)	N.A.	N.E.	0.0109
PATIENTS WITH MALIGNANCY CTC GRADE >=3	96	0	N.E.	49	2 ( 4.1)	N.A.	N.E.	0.0109
PATIENTS WITH SERIOUS MALIGNANCY	96	0	N.E.	49	2 ( 4.1)	N.A.	N.E.	0.0109

Time to first malignancy is defined as the time from first dose to the earliest malignancy event occurrence date. Subjects without malignancy will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfmmal-ebr700.sas

21JUN2022:07:07:21

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Premalignancy  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY PREMALIGNANCY	96	1 ( 1.0)	N.A.	49	1 ( 2.0)	N.A.	0.28 (0.02, 5.02)	0.3645
PATIENTS WITH PREMALIGNANCY CTC GRADE >=3	96	0	N.E.	49	0	N.E.	N.E.	N.E.
PATIENTS WITH SERIOUS PREMALIGNANCY	96	0	N.E.	49	0	N.E.	N.E.	N.E.

Time to first premalignancy is defined as the time from first dose the earliest premalignancy event occurrence date. Subjects without premalignancy will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfmmal-ebr700.sas

21JUN2022:07:07:27

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Thromboembolic and thrombophlebitis Events  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY THROMBOEMBOLIC AND THROMBOPHLEBITIS EVENTS	96	0	N.E.	49	0	N.E.	N.E.	N.E.
PATIENTS WITH THROMBOEMBOLIC AND THROMBOPHLEBITIS EVENTS CTC GRADE >=3	96	0	N.E.	49	0	N.E.	N.E.	N.E.
PATIENTS WITH SERIOUS THROMBOEMBOLIC AND THROMBOPHLEBITIS EVENTS	96	0	N.E.	49	0	N.E.	N.E.	N.E.

Time to first Thromboembolic and Thrombophlebitis events is defined as the time from first dose to the earliest event occurrence date. Subjects without Thromboembolic and Thrombophlebitis events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

Thromboembolic and Thrombophlebitis events are defined in terms of AE in the SMQ "Embolic and Thrombotic Events and Thrombophlebitis".

HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-T-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfmmal-ebr700.sas

21JUN2022:07:07:42

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Bone pain or bone pain-like event  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY BONE PAIN OR BONE LIKE EVENTS	96	65 ( 67.7)	1.28 (0.13, 10.02)	49	15 ( 30.6)	N.A.	3.16 (1.80, 5.56)	<0.0001
PATIENTS WITH BONE PAIN OR BONE LIKE EVENTS CTC GRADE >=3	96	3 ( 3.1)	N.A.	49	0	N.E.	N.E.	0.2170
PATIENTS WITH SERIOUS BONE PAIN OR BONE LIKE EVENTS	96	0	N.E.	49	0	N.E.	N.E.	N.E.

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain

HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NTD-T-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfmmal-ebr700.sas

21JUN2022:07:07:50

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Hypertension  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY HYPERTENSION EVENTS	96	19 ( 19.8)	N.A.	49	1 ( 2.0)	N.A.	9.32 (1.25, 69.66)	0.0079
PATIENTS WITH HYPERTENSION EVENTS CTC GRADE >=3	96	3 ( 3.1)	N.A.	49	0	N.E.	N.E.	0.2161
PATIENTS WITH SERIOUS HYPERTENSION EVENTS	96	0	N.E.	49	0	N.E.	N.E.	N.E.

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

Hypertension events are defined in terms of PT=Hypertension

HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfmmal-ebr700.sas

21JUN2022:07:07:35

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Extramedullary Haemopoiesis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY EXTRAMEDULLARY HAEMOPOIESIS EVENTS	96	7 ( 7.3)	N.A.	49	1 ( 2.0)	N.A.	2.99 (0.36, 24.56)	0.2845
PATIENTS WITH EXTRAMEDULLARY HAEMOPOIESIS EVENTS CTC GRADE >=3	96	0	N.E.	49	1 ( 2.0)	N.A.	N.E.	0.0833
PATIENTS WITH SERIOUS EXTRAMEDULLARY HAEMOPOIESIS EVENTS	96	0	N.E.	49	1 ( 2.0)	N.A.	N.E.	0.0833

Time to first extramedullary haemopoiesis event is defined as the time from first dose to the earliest event occurrence date. Subjects without extramedullary haemopoiesis events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

Extramedullary haemopoiesis events are defined in terms of PT=Extramedullary haemopoiesis or Chest wall mass  
HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfe-ebr700.sas

09FEB2023:08:55:36

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Traumatic Fracture  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY TRAUMATIC FRACTURE EVENTS	96	8 ( 8.3)	N.A.	49	1 ( 2.0)	N.A.	2.75 (0.33, 22.55)	0.3265
PATIENTS WITH TRAUMATIC FRACTURE EVENTS CTC GRADE >=3	96	4 ( 4.2)	N.A. (28.81, N.A.)	49	1 ( 2.0)	N.A.	1.27 (0.13, 12.30)	0.8343
PATIENTS WITH SERIOUS TRAUMATIC FRACTURE EVENTS	96	5 ( 5.2)	N.A.	49	1 ( 2.0)	N.A.	1.76 (0.20, 15.77)	0.6102

Time to first traumatic fracture event is defined as the time from first dose to the earliest event occurrence date. Subjects without traumatic fracture events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

Traumatic fracture events are defined in terms of PT=Traumatic fracture

HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE.

(2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3).

Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfe-ebr700.sas

09FEB2023:08:55:43

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Time to First Leg Ulcer  
Safety Population

	Luspatercept			Placebo			Luspatercept vs Placebo	
	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	N	Patients with Event n (%)	KME [95% CI] (mon) (1)	HR [95% CI] (2)	P-value (3)
PATIENTS WITH ANY LEG ULCER EVENTS	96	5 ( 5.2)	N.A.	49	2 ( 4.1)	N.A.	1.10 (0.21, 5.72)	0.9121
PATIENTS WITH LEG ULCER EVENTS CTC GRADE >=3	96	0	N.E.	49	0	N.E.	N.E.	N.E.
PATIENTS WITH SERIOUS LEG ULCER EVENTS	96	0	N.E.	49	0	N.E.	N.E.	N.E.

Time to first leg ulcer event is defined as the time from first dose to the earliest event occurrence date. Subjects without leg ulcer events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Leg ulcer events are defined in terms of LLT=Ankle ulcer, Foot ulcer, Infected skin ulcer, or Leg ulcer HR=hazard ratio; KME=Kaplan-Meier estimate. (1) KME of median time to first AE. (2) Stratified Cox proportional hazard model with treatment as only covariate and stratified by baseline Hb category (>= 8.5 g/dL, < 8.5g/dL) and baseline NIDT-PRO T/W score category (>=3 vs. <3). Hazard Ratio is Luspatercept over Placebo. (3) Stratified Log-rank Test to compare Luspatercept and Placebo. Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfe-ebr700.sas 09FEB2023:08:55:50

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

Protocol: ACE-536-B-THAL-002

Page 1 of 1

TEAE Leading to Study Drug Discontinuation by Treatment, System Organ Class and Preferred Term  
Safety Population

System Organ Class (%) Preferred Term (%)	Luspatercept N = 96	Placebo N = 49
SUBJECT WITH AT LEAST ONE TEAE LEADING TO STUDY DRUG DISCONTINUATION	3 ( 3.1)	4 ( 8.2)
Blood and lymphatic system disorders	2 ( 2.1)	0
Extramedullary haemopoiesis	1 ( 1.0)	0
Haemolytic anaemia	1 ( 1.0)	0
General disorders and administration site conditions	0	2 ( 4.1)
Asthenia	0	1 ( 2.0)
Fatigue	0	1 ( 2.0)
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	0	2 ( 4.1)
Diffuse large B-cell lymphoma	0	1 ( 2.0)
Hepatocellular carcinoma	0	1 ( 2.0)
Musculoskeletal and connective tissue disorders	1 ( 1.0)	0
Arthralgia	1 ( 1.0)	0
Lupus-like syndrome	1 ( 1.0)	0
Nervous system disorders	1 ( 1.0)	0
Spinal cord compression	1 ( 1.0)	0
Respiratory, thoracic and mediastinal disorders	1 ( 1.0)	0
Pulmonary arterial hypertension	1 ( 1.0)	0
Vascular disorders	0	1 ( 2.0)
Hypotension	0	1 ( 2.0)

Treatment-Emergent Adverse Events (TEAE), is defined as any AE or SAE occurring or worsening on or after the day of the first dose of the IP through 63 days after the last dose of IP. In addition, any AE with an onset date beyond this timeframe and that is assessed by the investigator as related to IP will be considered a TEAE.

Adverse events are coded using MedDRA Version 23.0 A subject is counted only once for multiple events within each SOC/PT regardless of relationship (causality).

System organ classes (SOCs) and preferred terms (PTs) are presented in descending frequency within the Luspatercept group.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-safcat-eb700.sas

22NOV2022:10:25:01

Protocol: ACE-536-B-THAL-002

Page 1 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
OVERALL	96	96 (100.0)	0.10 ( 0.07, 0.13)	49	48 ( 98.0)	0.76 ( 0.46, 0.89)	1.54 ( 1.08, 2.19) 0.0199	
REGION								0.4465
NORTH AMERICA & EUROPE	60	60 (100.0)	0.10 ( 0.03, 0.10)	30	30 (100.0)	0.72 ( 0.46, 0.82)	1.53 ( 0.98, 2.39) 0.0924	
MIDDLE EAST	9	9 (100.0)	0.36 ( 0.03, 0.89)	8	8 (100.0)	0.05 ( 0.03, 1.74)	1.10 ( 0.40, 3.03) 0.7577	
ASIA-PACIFIC	27	27 (100.0)	0.13 ( 0.07, 0.89)	11	10 ( 90.9)	2.04 ( 0.10, 5.13)	2.10 ( 0.98, 4.51) 0.0518	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-eb700.sas

30MAY2022:07:56:47

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

Protocol: ACE-536-B-THAL-002

Page 2 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo		Test for Interaction P-value (4) (5)
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)		
AGE									0.2252
<=32	35	35 (100.0)	0.36 ( 0.07, 1.41)	12	11 ( 91.7)	0.74 ( 0.03, 3.52)	1.20 ( 0.61, 2.37)		
>32	61	61 (100.0)	0.10 ( 0.07, 0.10)	37	37 (100.0)	0.76 ( 0.46, 1.05)	2.16 ( 1.41, 3.31)	0.6086 0.0004	
SPLENECTOMY									0.3800
YES	34	34 (100.0)	0.10 ( 0.03, 0.49)	26	25 ( 96.2)	0.67 ( 0.07, 1.05)	1.25 ( 0.74, 2.10)	0.4287	
NO	62	62 (100.0)	0.10 ( 0.07, 0.10)	23	23 (100.0)	0.76 ( 0.26, 1.12)	1.79 ( 1.10, 2.94)	0.0233	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-eb700.sas

30MAY2022:07:56:47

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

Protocol: ACE-536-B-THAL-002

Page 3 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo		Test for Interaction P-value (4) (5)
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)		
SEX									0.2753
MALE	40	40 (100.0)	0.13 ( 0.07, 0.39)	23	23 (100.0)	0.66 ( 0.07, 1.05)	1.26 ( 0.75, 2.13)		
FEMALE	56	56 (100.0)	0.10 ( 0.07, 0.10)	26	25 ( 96.2)	0.77 ( 0.13, 1.31)	1.83 ( 1.13, 2.95)	0.3858 0.0167	
BETA-THALASSEMIA DIAGNOSIS									0.2173
BETA-THALASSEMIA	63	63 (100.0)	0.10 ( 0.07, 0.10)	34	34 (100.0)	0.72 ( 0.13, 0.82)	1.34 ( 0.88, 2.06)	0.2237	
HEMOGLOBIN E/BETA-THALASSEMIA	28	28 (100.0)	0.13 ( 0.07, 0.79)	11	10 ( 90.9)	2.04 ( 0.10, 5.13)	2.24 ( 1.04, 4.82)	0.0344	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-eb700.sas

30MAY2022:07:56:47

Protocol: ACE-536-B-THAL-002

Page 4 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo		Test for Interaction P-value (4) (5)
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)		
BETA-THALASSEMIA GENE MUTATION									0.7781
[B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	69 (100.0)	0.10 (0.07, 0.20)	33	32 ( 97.0)	0.82 (0.13, 1.61)	1.50 (0.98, 2.30)		
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	6 (100.0)	0.05 (0.03, N.A.)	4	4 (100.0)	0.30 (0.03, N.A.)	2.57 (0.49, 13.31)		
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	21 (100.0)	0.10 (0.03, 0.13)	12	12 (100.0)	0.74 (0.03, 1.05)	2.14 (1.00, 4.59)		
BASELINE HEMOGLOBIN <8.5 G/DL	55	55 (100.0)	0.10 (0.07, 0.43)	29	28 ( 96.6)	0.85 (0.26, 1.74)	1.71 (1.08, 2.73)		0.3186
>=8.5 G/DL	41	41 (100.0)	0.10 (0.07, 0.13)	20	20 (100.0)	0.72 (0.03, 0.79)	1.30 (0.75, 2.25)		

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-ebr700.sas

30MAY2022:07:56:47

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

Protocol: ACE-536-B-THAL-002

Page 5 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE NTD-T-PRO T/W SCORE								0.5627
<3 POINTS	30	30 (100.0)	0.10 ( 0.07, 0.43)	14	13 ( 92.9)	0.13 ( 0.03, 3.52)	1.60 ( 0.81, 3.19)	
>=3 POINTS	66	66 (100.0)	0.10 ( 0.07, 0.10)	35	35 (100.0)	0.79 ( 0.66, 1.05)	1.41 ( 0.93, 2.15)	0.1447

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-ubr700.sas

30MAY2022:07:56:47

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

Protocol: ACE-536-B-THAL-002

Page 6 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE LIVER IRON								0.9212
<3 MG/G DRY WEIGHT	35	35 (100.0)	0.10 ( 0.07, 0.33)	17	17 (100.0)	0.72 ( 0.13, 0.82)	1.21 ( 0.67, 2.19)	
>=3 TO <=5 MG/G DRY WEIGHT	23	23 (100.0)	0.10 ( 0.07, 0.10)	13	13 (100.0)	0.79 ( 0.07, 1.61)	0.5625 1.60 ( 0.79, 3.26)	
>5 TO <=7 MG/G DRY WEIGHT	12	12 (100.0)	0.10 ( 0.03, 0.89)	3	3 (100.0)	1.05 ( 0.56, N.A.)	0.2217 1.99 ( 0.54, 7.31)	
>7 TO <=15 MG/G DRY WEIGHT	14	14 (100.0)	0.11 ( 0.07, 0.89)	9	8 ( 88.9)	0.89 ( 0.03, 2.14)	0.3086 1.60 ( 0.66, 3.87)	
>15 MG/G DRY WEIGHT	11	11 (100.0)	0.10 ( 0.03, 2.17)	5	5 (100.0)	0.10 ( 0.03, N.A.)	0.2895 1.43 ( 0.45, 4.53)	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-ubr700.sas

30MAY2022:07:56:47

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

Protocol: ACE-536-B-THAL-002

Page 7 of 7

Time to first Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE 6MWT <=450 METERS	51	51 (100.0)	0.10 ( 0.07, 0.13)	31	30 ( 96.8)	0.85 ( 0.13, 1.91)	1.90 ( 1.20, 3.02)	0.2068
>450 METERS	45	45 (100.0)	0.10 ( 0.07, 0.13)	18	18 (100.0)	0.64 ( 0.07, 0.79)	0.0065 ( 1.26, 2.23)	0.4791
BASELINE FACIT-F FS SCORE <43 POINTS	64	64 (100.0)	0.10 ( 0.07, 0.13)	32	32 (100.0)	0.79 ( 0.56, 1.05)	1.43 ( 0.93, 2.21)	0.6729
>=43 POINTS	32	32 (100.0)	0.10 ( 0.07, 0.20)	16	15 ( 93.8)	0.20 ( 0.07, 2.20)	0.1231 ( 1.63, 3.08)	0.1238

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae-ebr700.sas

30MAY2022:07:56:47

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

Protocol: ACE-536-B-THAL-002

Page 1 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
OVERALL	96	11 ( 11.5)	N.A.	49	12 ( 24.5)	N.A. (18.00, N.A.)	0.29 ( 0.12, 0.68) 0.0026	
REGION								0.3649
NORTH AMERICA & EUROPE	60	6 ( 10.0)	N.A.	30	4 ( 13.3)	N.A.	0.46 ( 0.12, 1.76) 0.2465	
MIDDLE EAST	9	1 ( 11.1)	N.A. ( 7.89, N.A.)	8	6 ( 75.0)	18.00 ( 3.58, N.A.)	0.07 (<0.01, 0.66) 0.0041	
ASIA-PACIFIC	27	4 ( 14.8)	N.A.	11	2 ( 18.2)	N.A. (16.62, N.A.)	0.80 ( 0.15, 4.37) 0.7944	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 2 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
AGE								0.8462
<=32	35	3 ( 8.6)	N.A.	12	3 ( 25.0)	19.75 (19.48, N.A.)	0.22 ( 0.04, 1.14)	
>32	61	8 ( 13.1)	N.A. (28.81, N.A.)	37	9 ( 24.3)	N.A. (16.39, N.A.)	0.32 ( 0.12, 0.88)	0.0202
SPLENECTOMY								0.0365**
YES	34	3 ( 8.8)	N.A. (28.81, N.A.)	26	10 ( 38.5)	19.48 (12.42, N.A.)	0.08 ( 0.02, 0.37)	
NO	62	8 ( 12.9)	N.A.	23	2 ( 8.7)	N.A. (18.00, N.A.)	1.18 ( 0.25, 5.62)	0.8315

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 3 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
SEX								0.9974
MALE	40	5 ( 12.5)	N.A.	23	5 ( 21.7)	N.A. (16.62, N.A.)	0.44 ( 0.13, 1.56)	
FEMALE	56	6 ( 10.7)	N.A. (28.81, N.A.)	26	7 ( 26.9)	N.A. (18.00, N.A.)	0.1939 0.26 ( 0.08, 0.81) 0.0127	
BETA-THALASSEMIA DIAGNOSIS								0.5626
BETA-THALASSEMIA	63	8 ( 12.7)	N.A.	34	7 ( 20.6)	19.48 (18.00, N.A.)	0.29 ( 0.10, 0.88) 0.0214	
HEMOGLOBIN E/BETA-THALASSEMIA	28	3 ( 10.7)	N.A.	11	2 ( 18.2)	N.A. (16.62, N.A.)	0.62 ( 0.10, 3.72) 0.5993	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ebr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 4 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	8 ( 11.6)	N.A. (28.81, N.A.)	33	9 ( 27.3)	N.A. (18.00, N.A.)	0.27 ( 0.10, 0.74)	>0.9999
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	0	N.E.	4	3 ( 75.0)	14.34 ( 0.56, N.A.)	N.E.	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	3 ( 14.3)	N.A.	12	0	N.E.	N.E.	
BASELINE HEMOGLOBIN <8.5 G/DL	55	5 ( 9.1)	N.A.	29	7 ( 24.1)	N.A. (16.62, N.A.)	0.26 ( 0.08, 0.84)	0.5846
>=8.5 G/DL	41	6 ( 14.6)	28.81 (28.81, N.A.)	20	5 ( 25.0)	19.48 (12.42, N.A.)	0.33 ( 0.09, 1.17)	0.0158 0.0732

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 5 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE NTD-T-PRO T/W SCORE								0.9517
<3 POINTS	30	2 ( 6.7)	N.A.	14	3 ( 21.4)	N.A. (18.00, N.A.)	0.23 ( 0.04, 1.38)	
>=3 POINTS	66	9 ( 13.6)	N.A. (28.81, N.A.)	35	9 ( 25.7)	N.A. (16.39, N.A.)	0.33 ( 0.12, 0.87)	0.0194

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 6 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE LIVER IRON								
<3 MG/G DRY WEIGHT	35	4 ( 11.4)	N.A.	17	1 ( 5.9)	N.A. (18.00, N.A.)	N.M.E.	N.M.E.
>=3 TO <=5 MG/G DRY WEIGHT	23	2 ( 8.7)	N.A. (28.81, N.A.)	13	2 ( 15.4)	N.A. ( 9.89, N.A.)	N.M.E.	
>5 TO <=7 MG/G DRY WEIGHT	12	3 ( 25.0)	N.A. (11.07, N.A.)	3	1 ( 33.3)	N.A. ( 0.56, N.A.)	N.M.E.	
>7 TO <=15 MG/G DRY WEIGHT	14	2 ( 14.3)	N.A.	9	5 ( 55.6)	16.39 ( 3.58, N.A.)	N.M.E.	
>15 MG/G DRY WEIGHT	11	0	N.E.	5	1 ( 20.0)	N.A. (12.29, N.A.)	N.M.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ebr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 7 of 7

Time to first Serious Adverse Events  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE 6MWT <=450 METERS	51	7 ( 13.7)	N.A. (28.81, N.A.)	31	10 ( 32.3)	19.75 (16.62, N.A.)	0.26 ( 0.09, 0.71) 0.0052	0.8098
>450 METERS	45	4 ( 8.9)	N.A.	18	2 ( 11.1)	N.A.	0.58 ( 0.10, 3.30) 0.5374	
BASELINE FACIT-F FS SCORE <43 POINTS	64	9 ( 14.1)	N.A. (28.81, N.A.)	32	10 ( 31.3)	18.00 (16.39, N.A.)	0.27 ( 0.10, 0.70) 0.0043	0.8227
>=43 POINTS	32	2 ( 6.3)	N.A.	16	2 ( 12.5)	N.A. (19.75, N.A.)	0.35 ( 0.05, 2.49) 0.2716	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ebr700.sas

04AUG2022:08:53:34

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

Protocol: ACE-536-B-THAL-002

Page 1 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
OVERALL	96	27 ( 28.1)	N.A.	49	12 ( 24.5)	N.A. (16.62, N.A.)	1.02 ( 0.51, 2.02) 0.9584	
REGION								0.3069
NORTH AMERICA & EUROPE	60	11 ( 18.3)	N.A.	30	4 ( 13.3)	N.A. (15.64, N.A.)	1.18 ( 0.37, 3.75) 0.7820	
MIDDLE EAST	9	4 ( 44.4)	N.A. ( 0.07, N.A.)	8	6 ( 75.0)	18.00 ( 3.58, N.A.)	0.58 ( 0.16, 2.15) 0.4120	
ASIA-PACIFIC	27	12 ( 44.4)	N.A. ( 9.03, N.A.)	11	2 ( 18.2)	N.A. (16.62, N.A.)	3.05 ( 0.68, 13.65) 0.1250	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

21JUN2022:07:04:43

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

Protocol: ACE-536-B-THAL-002

Page 2 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo		Test for Interaction P-value (4) (5)
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)		
AGE									0.5789
<=32	35	7 ( 20.0)	N.A.	12	3 ( 25.0)	19.75 (19.48, N.A.)	0.75 ( 0.19, 2.95)		
>32	61	20 ( 32.8)	N.A.	37	9 ( 24.3)	N.A. (16.39, N.A.)	1.17 ( 0.53, 2.60)		
SPLENECTOMY									0.0293**
YES	34	8 ( 23.5)	N.A.	26	10 ( 38.5)	16.62 (12.42, N.A.)	0.48 ( 0.18, 1.25)		
NO	62	19 ( 30.6)	N.A.	23	2 ( 8.7)	N.A. (18.00, N.A.)	3.38 ( 0.79, 14.54)		

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ebr700.sas

21JUN2022:07:04:43

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

Protocol: ACE-536-B-THAL-002

Page 3 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
SEX								0.6844
MALE	40	12 ( 30.0)	N.A.	23	6 ( 26.1)	16.62 (15.64, N.A.)	1.03 ( 0.38, 2.77)	
FEMALE	56	15 ( 26.8)	N.A.	26	6 ( 23.1)	N.A. (18.00, N.A.)	0.9578 1.13 ( 0.44, 2.93)	
BETA-THALASSEMIA DIAGNOSIS								0.1540
BETA-THALASSEMIA	63	14 ( 22.2)	N.A.	34	7 ( 20.6)	19.48 (15.64, N.A.)	0.84 ( 0.33, 2.14)	
HEMOGLOBIN E/BETA-THALASSEMIA	28	12 ( 42.9)	N.A. (12.88, N.A.)	11	2 ( 18.2)	N.A. (16.62, N.A.)	0.7190 3.03 ( 0.68, 13.56)	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ebr700.sas

21JUN2022:07:04:43

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

Protocol: ACE-536-B-THAL-002

Page 4 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	21 ( 30.4)	N.A.	33	8 ( 24.2)	N.A. (18.00, N.A.)	1.22 ( 0.54, 2.77)	0.1782
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	1 ( 16.7)	N.A. ( 3.94, N.A.)	4	3 ( 75.0)	14.34 ( 0.56, N.A.)	0.16 ( 0.02, 1.60)	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	5 ( 23.8)	N.A. (15.57, N.A.)	12	1 ( 8.3)	N.A. (15.64, N.A.)	2.21 ( 0.25, 19.72)	0.4681
BASELINE HEMOGLOBIN <8.5 G/DL	55	12 ( 21.8)	N.A.	29	6 ( 20.7)	N.A. (16.62, N.A.)	0.82 ( 0.31, 2.22)	0.6812
>=8.5 G/DL	41	15 ( 36.6)	N.A. ( 7.89, N.A.)	20	6 ( 30.0)	19.48 (12.42, N.A.)	0.7025 1.25 ( 0.48, 3.23)	0.6560

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

21JUN2022:07:04:43

Protocol: ACE-536-B-THAL-002

Page 5 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE NTD-T-PRO T/W SCORE <3 POINTS	30	9 ( 30.0)	N.A. (15.57, N.A.)	14	4 ( 28.6)	N.A. (18.00, N.A.)	1.09 ( 0.34, 3.57)	0.8433
>=3 POINTS	66	18 ( 27.3)	N.A.	35	8 ( 22.9)	N.A. (16.39, N.A.)	1.05 ( 0.45, 2.43)	0.9134

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

21JUN2022:07:04:43

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

Protocol: ACE-536-B-THAL-002

Page 6 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	9 ( 25.7)	N.A.	17	1 ( 5.9)	N.A. (18.00, N.A.)	4.14 ( 0.52, 32.89)	0.2180
>=3 TO <=5 MG/G DRY WEIGHT	23	8 ( 34.8)	N.A. ( 7.16, N.A.)	13	1 ( 7.7)	N.A.	0.1451 4.97 ( 0.62, 39.91)	
>5 TO <=7 MG/G DRY WEIGHT	12	3 ( 25.0)	N.A. ( 0.13, N.A.)	3	1 ( 33.3)	N.A. ( 0.56, N.A.)	0.0941 0.67 ( 0.07, 6.64)	
>7 TO <=15 MG/G DRY WEIGHT	14	4 ( 28.6)	N.A. ( 7.89, N.A.)	9	5 ( 55.6)	16.39 ( 3.58, N.A.)	0.7307 0.40 ( 0.11, 1.53)	
>15 MG/G DRY WEIGHT	11	3 ( 27.3)	N.A. ( 8.87, N.A.)	5	2 ( 40.0)	N.A. (12.29, N.A.)	0.1691 0.77 ( 0.13, 4.67)	0.7766

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ibr700.sas

21JUN2022:07:04:43

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

Protocol: ACE-536-B-THAL-002

Page 7 of 7

Time to first Adverse Events CTC Grade >=3  
Subgroups Analysis  
Safety Population

	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4) (5)
BASELINE 6MWT <=450 METERS	51	19 ( 37.3)	N.A. (15.44, N.A.)	31	10 ( 32.3)	19.75 (16.39, N.A.)	1.17 ( 0.54, 2.52) 0.6999	0.9260
>450 METERS	45	8 ( 17.8)	N.A.	18	2 ( 11.1)	N.A.	1.30 ( 0.27, 6.20) 0.7400	
BASELINE FACIT-F FS SCORE <43 POINTS	64	21 ( 32.8)	N.A.	32	9 ( 28.1)	19.48 (16.39, N.A.)	1.01 ( 0.46, 2.22) 0.9852	0.8891
>=43 POINTS	32	6 ( 18.8)	N.A.	16	3 ( 18.8)	N.A. (15.64, N.A.)	0.95 ( 0.24, 3.82) 0.9431	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first).

HR=hazard ratio; (1) KME=Kaplan-Meier estimate.

(2) Unstratified Cox proportional hazard model. HR is Luspatercept over Placebo (3) Unstratified log-rank test.

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

(5) A p-value of <0.05 needs to be indicated by 2 asterisks (proof of different effects).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-tfsubae2-ebr700.sas

21JUN2022:07:04:43

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

Protocol: ACE-536-B-THAL-002

Page 1 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	76 ( 79.2)	0.72 ( 0.13, 2.07)	49	28 ( 57.1)	8.51 ( 1.74, N.A.)	1.84 ( 1.19, 2.84) 0.0066	
REGION								0.0437
NORTH AMERICA & EUROPE	60	48 ( 80.0)	0.30 ( 0.10, 2.07)	30	14 ( 46.7)	N.A. ( 1.38, N.A.)	2.49 ( 1.37, 4.53) 0.0025	
MIDDLE EAST	9	7 ( 77.8)	2.17 ( 0.03, N.A.)	8	8 (100.0)	0.77 ( 0.03, 8.51)	0.57 ( 0.21, 1.59) 0.2816	
ASIA-PACIFIC	27	21 ( 77.8)	0.79 ( 0.10, 10.91)	11	6 ( 54.5)	13.90 ( 0.13, N.A.)	2.09 ( 0.84, 5.21) 0.1067	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-eb700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 2 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								
<=32	35	24 ( 68.6)	3.35 ( 0.79, 15.57)	12	9 ( 75.0)	4.85 ( 0.07, 16.43)	0.84 ( 0.39, 1.83)	0.0173
>32	61	52 ( 85.2)	0.13 ( 0.10, 0.76)	37	19 ( 51.4)	8.51 ( 1.74, N.A.)	0.6583 2.73 ( 1.61, 4.63) 0.0001	
SPLENECTOMY								
YES	34	24 ( 70.6)	2.35 ( 0.33, 14.59)	26	15 ( 57.7)	1.91 ( 0.95, N.A.)	1.26 ( 0.66, 2.42)	0.1796
NO	62	52 ( 83.9)	0.13 ( 0.10, 1.45)	23	13 ( 56.5)	10.32 ( 2.20, N.A.)	0.4963 2.27 ( 1.23, 4.19) 0.0081	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 3 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								0.6273
MALE	40	28 ( 70.0)	1.81 ( 0.13, 11.50)	23	12 ( 52.2)	8.51 ( 1.38, N.A.)	1.63 ( 0.82, 3.21)	0.1698
FEMALE	56	48 ( 85.7)	0.30 ( 0.10, 1.54)	26	16 ( 61.5)	8.66 ( 0.82, N.A.)	1.99 ( 1.12, 3.51)	0.0175
BETA-THALASSEMIA DIAGNOSIS								0.5033
BETA-THALASSEMIA	63	49 ( 77.8)	0.49 ( 0.10, 2.17)	34	20 ( 58.8)	5.80 ( 1.05, N.A.)	1.66 ( 0.98, 2.80)	0.0638
HEMOGLOBIN E/BETA-THALASSEMIA	28	23 ( 82.1)	0.77 ( 0.10, 2.53)	11	6 ( 54.5)	13.90 ( 0.13, N.A.)	2.36 ( 0.95, 5.84)	0.0553

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 4 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION								0.6649
[B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	52 ( 75.4)	1.45 ( 0.13, 3.35)	33	19 ( 57.6)	10.32 ( 1.81, N.A.)	1.60 ( 0.94, 2.71)	
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	5 ( 83.3)	0.53 ( 0.03, N.A.)	4	2 ( 50.0)	1.74 ( 1.38, N.A.)	0.0837 2.35 ( 0.45, 12.35)	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	19 ( 90.5)	0.13 ( 0.10, 1.51)	12	7 ( 58.3)	5.80 ( 0.82, N.A.)	0.3108 2.86 ( 1.19, 6.87)	
BASELINE HEMOGLOBIN <8.5 G/DL	55	43 ( 78.2)	1.51 ( 0.13, 5.36)	29	14 ( 48.3)	16.43 ( 1.81, N.A.)	2.02 ( 1.10, 3.70)	0.4851
>=8.5 G/DL	41	33 ( 80.5)	0.13 ( 0.10, 1.54)	20	14 ( 70.0)	2.84 ( 0.82, 8.51)	0.0212 1.70 ( 0.91, 3.20)	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 5 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTD-T-PRO T/W SCORE								0.7564
<3 POINTS	30	22 ( 73.3)	0.46 ( 0.10, 5.36)	14	8 ( 57.1)	13.90 ( 0.07, N.A.)	1.59 ( 0.70, 3.58)	0.2665
>=3 POINTS	66	54 ( 81.8)	0.72 ( 0.13, 2.17)	35	20 ( 57.1)	7.00 ( 1.74, N.A.)	1.94 ( 1.15, 3.25)	0.0125

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 6 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	25 ( 71.4)	2.07 ( 0.13, 11.89)	17	11 ( 64.7)	2.20 ( 0.72, N.A.)	1.13 ( 0.55, 2.33)	0.4374
>=3 TO <=5 MG/G DRY WEIGHT	23	20 ( 87.0)	0.10 ( 0.10, 1.45)	13	7 ( 53.8)	7.00 ( 0.95, N.A.)	0.7483 2.66 ( 1.11, 6.37)	
>5 TO <=7 MG/G DRY WEIGHT	12	9 ( 75.0)	0.31 ( 0.03, 22.11)	3	0	N.E.	0.0256 N.E.	
>7 TO <=15 MG/G DRY WEIGHT	14	11 ( 78.6)	0.62 ( 0.07, 11.50)	9	6 ( 66.7)	6.21 ( 0.07, N.A.)	1.54 ( 0.57, 4.18)	
>15 MG/G DRY WEIGHT	11	10 ( 90.9)	2.04 ( 0.33, 2.53)	5	2 ( 40.0)	N.A. ( 1.38, N.A.)	0.4086 5.05 ( 1.05, 24.40)	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 7 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT								
<=450 METERS	51	42 ( 82.4)	0.76 ( 0.13, 2.07)	31	16 ( 51.6)	13.90 ( 1.91, N.A.)	2.32 ( 1.30, 4.15)	0.1673
>450 METERS	45	34 ( 75.6)	0.69 ( 0.10, 5.36)	18	12 ( 66.7)	1.81 ( 0.82, N.A.)	0.0036 1.27 ( 0.65, 2.46) 0.5281	
BASELINE FACIT-F FS SCORE								
<43 POINTS	64	54 ( 84.4)	0.72 ( 0.13, 2.07)	32	19 ( 59.4)	6.21 ( 1.31, N.A.)	1.91 ( 1.13, 3.24)	0.9557
>=43 POINTS	32	22 ( 68.8)	0.89 ( 0.10, 14.59)	16	8 ( 50.0)	16.43 ( 0.82, N.A.)	0.0159 1.71 ( 0.76, 3.85) 0.1952	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 8 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	35 ( 36.5)	N.A.	49	3 ( 6.1)	N.A.	7.26 ( 2.23, 23.63) 0.0001	
REGION								0.9762
NORTH AMERICA & EUROPE	60	24 ( 40.0)	N.A. ( 2.92, N.A.)	30	2 ( 6.7)	N.A.	7.30 ( 1.72, 30.95) 0.0016	
MIDDLE EAST	9	0	N.E.	8	0	N.E.	N.E.	
ASIA-PACIFIC	27	11 ( 40.7)	N.A. ( 0.13, N.A.)	11	1 ( 9.1)	N.A.	5.77 ( 0.74, 44.76) 0.0580	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 9 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								
<=32	35	7 ( 20.0)	N.A.	12	1 ( 8.3)	N.A.	2.58 ( 0.32, 20.98)	0.2478
>32	61	28 ( 45.9)	N.A. ( 0.13, N.A.)	37	2 ( 5.4)	N.A.	0.3581 11.14 ( 2.65, 46.83) <0.0001	
SPLENECTOMY								
YES	34	6 ( 17.6)	N.A.	26	1 ( 3.8)	N.A.	4.64 ( 0.56, 38.56)	0.7725
NO	62	29 ( 46.8)	N.A. ( 0.13, N.A.)	23	2 ( 8.7)	N.A.	0.1186 7.23 ( 1.72, 30.34) 0.0015	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 10 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								
MALE	40	13 ( 32.5)	N.A.	23	1 ( 4.3)	N.A.	8.44 ( 1.10, 64.61)	0.8026
FEMALE	56	22 ( 39.3)	N.A. ( 2.92, N.A.)	26	2 ( 7.7)	N.A.	0.0139 6.45 ( 1.52, 27.46) 0.0037	
BETA-THALASSEMIA DIAGNOSIS								
BETA-THALASSEMIA	63	23 ( 36.5)	N.A. (14.59, N.A.)	34	2 ( 5.9)	N.A.	7.43 ( 1.75, 31.58)	0.7931
HEMOGLOBIN E/BETA-THALASSEMIA	28	11 ( 39.3)	N.A. ( 0.13, N.A.)	11	1 ( 9.1)	N.A.	0.0014 5.50 ( 0.71, 42.67) 0.0669	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 11 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	24 ( 34.8)	N.A.	33	2 ( 6.1)	N.A.	6.89 ( 1.63, 29.19) 0.0023	0.9952
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	1 ( 16.7)	N.A. ( 0.07, N.A.)	4	0	N.E.	N.E.	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	10 ( 47.6)	N.A. ( 0.10, N.A.)	12	1 ( 8.3)	N.A.	7.78 ( 0.99, 61.03) 0.0214	
BASELINE HEMOGLOBIN <8.5 G/DL	55	18 ( 32.7)	N.A.	29	1 ( 3.4)	N.A.	10.96 ( 1.46, 82.10) 0.0034	0.5481
>=8.5 G/DL	41	17 ( 41.5)	N.A. ( 0.13, N.A.)	20	2 ( 10.0)	N.A.	5.43 ( 1.25, 23.53) 0.0115	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 12 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTD-T-PRO T/W SCORE <3 POINTS	30	12 ( 40.0)	N.A. ( 0.13, N.A.)	14	1 ( 7.1)	N.A.	7.00 ( 0.91, 53.91) 0.0295	0.9941
>=3 POINTS	66	23 ( 34.8)	N.A.	35	2 ( 5.7)	N.A.	7.20 ( 1.70, 30.57) 0.0018	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 13 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	13 ( 37.1)	N.A. ( 2.92, N.A.)	17	2 ( 11.8)	N.A.	3.85 ( 0.87, 17.06) 0.0570	0.9763
>=3 TO <=5 MG/G DRY WEIGHT	23	13 ( 56.5)	1.45 ( 0.10, N.A.)	13	0	N.E.	N.E.	
>5 TO <=7 MG/G DRY WEIGHT	12	6 ( 50.0)	N.A. ( 0.03, N.A.)	3	0	N.E.	N.E.	
>7 TO <=15 MG/G DRY WEIGHT	14	2 ( 14.3)	N.A.	9	1 ( 11.1)	N.A. ( 1.31, N.A.)	1.39 ( 0.13, 15.33) 0.8007	
>15 MG/G DRY WEIGHT	11	1 ( 9.1)	N.A.	5	0	N.E.	N.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 14 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Bone Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT <=450 METERS	51	16 ( 31.4)	N.A.	31	1 ( 3.2)	N.A.	11.40 ( 1.51, 85.99) 0.0028	0.4817
>450 METERS	45	19 ( 42.2)	N.A. ( 0.13, N.A.)	18	2 ( 11.1)	N.A.	4.78 ( 1.11, 20.56) 0.0206	
BASELINE FACIT-F FS SCORE <43 POINTS	64	21 ( 32.8)	N.A.	32	1 ( 3.1)	N.A.	12.62 ( 1.70, 93.84) 0.0014	0.4079
>=43 POINTS	32	14 ( 43.8)	N.A. ( 0.13, N.A.)	16	2 ( 12.5)	N.A.	4.36 ( 0.99, 19.22) 0.0342	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 15 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	27 ( 28.1)	N.A. (28.91, N.A.)	49	5 ( 10.2)	N.A.	2.54 ( 0.97, 6.64) 0.0487	
REGION								0.5522
NORTH AMERICA & EUROPE	60	20 ( 33.3)	N.A. (22.90, N.A.)	30	2 ( 6.7)	N.A.	4.55 ( 1.05, 19.71) 0.0261	
MIDDLE EAST	9	3 ( 33.3)	N.A. ( 1.12, N.A.)	8	2 ( 25.0)	N.A. ( 2.07, N.A.)	1.48 ( 0.25, 8.86) 0.6687	
ASIA-PACIFIC	27	4 ( 14.8)	N.A.	11	1 ( 9.1)	N.A.	1.57 ( 0.17, 14.06) 0.6862	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 16 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								0.9898
<=32	35	9 ( 25.7)	N.A.	12	0	N.E.	N.E.	
>32	61	18 ( 29.5)	N.A. (28.91, N.A.)	37	5 ( 13.5)	N.A.	1.95 ( 0.72, 5.31) 0.1842	
SPLENECTOMY								0.4205
YES	34	7 ( 20.6)	N.A.	26	3 ( 11.5)	N.A.	1.81 ( 0.47, 7.02) 0.3845	
NO	62	20 ( 32.3)	28.91 (28.91, N.A.)	23	2 ( 8.7)	N.A. (17.74, N.A.)	3.48 ( 0.81, 14.96) 0.0748	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 17 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								0.7686
MALE	40	9 ( 22.5)	N.A.	23	2 ( 8.7)	N.A.	2.25 ( 0.48, 10.58)	0.2931
FEMALE	56	18 ( 32.1)	28.91 (24.25, N.A.)	26	3 ( 11.5)	N.A.	2.78 ( 0.81, 9.49)	0.0894
BETA-THALASSEMIA DIAGNOSIS								0.8413
BETA-THALASSEMIA	63	21 ( 33.3)	N.A. (22.90, N.A.)	34	4 ( 11.8)	N.A. (17.74, N.A.)	2.44 ( 0.83, 7.22)	0.0954
HEMOGLOBIN E/BETA-THALASSEMIA	28	5 ( 17.9)	N.A.	11	1 ( 9.1)	N.A.	1.96 ( 0.23, 16.81)	0.5319

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 18 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	17 ( 24.6)	N.A.	33	3 ( 9.1)	N.A.	2.59 ( 0.76, 8.86)	0.9225
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	2 ( 33.3)	N.A. ( 0.03, N.A.)	4	0	N.E.	N.E.	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	8 ( 38.1)	28.91 (12.16, N.A.)	12	2 ( 16.7)	N.A. ( 7.00, N.A.)	1.67 ( 0.33, 8.32)	0.5293
BASELINE HEMOGLOBIN <8.5 G/DL	55	14 ( 25.5)	N.A. (28.91, N.A.)	29	1 ( 3.4)	N.A.	6.25 ( 0.81, 47.98)	0.2288
>=8.5 G/DL	41	13 ( 31.7)	N.A. (22.90, N.A.)	20	4 ( 20.0)	N.A. (17.74, N.A.)	1.56 ( 0.51, 4.79)	0.4362

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 19 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTDI-PRO T/W SCORE								0.3093
<3 POINTS	30	6 ( 20.0)	N.A. (28.91, N.A.)	14	2 ( 14.3)	N.A. (17.74, N.A.)	1.11 ( 0.21, 5.76) 0.9037	
>=3 POINTS	66	21 ( 31.8)	N.A. (22.90, N.A.)	35	3 ( 8.6)	N.A.	3.59 ( 1.07, 12.07) 0.0276	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 20 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	10 ( 28.6)	28.91 (28.91, N.A.)	17	2 ( 11.8)	N.A. (17.74, N.A.)	1.68 ( 0.36, 7.96)	0.9996
>=3 TO <=5 MG/G DRY WEIGHT	23	7 ( 30.4)	N.A. (22.90, N.A.)	13	2 ( 15.4)	N.A. ( 7.00, N.A.)	0.5057 ( 0.36, 8.60)	
>5 TO <=7 MG/G DRY WEIGHT	12	2 ( 16.7)	N.A. ( 1.45, N.A.)	3	0	N.E.	1.77 ( 0.4760, N.E.)	
>7 TO <=15 MG/G DRY WEIGHT	14	4 ( 28.6)	N.A. ( 0.07, N.A.)	9	1 ( 11.1)	N.A. ( 2.07, N.A.)	3.03 ( 0.34, 27.21)	
>15 MG/G DRY WEIGHT	11	4 ( 36.4)	N.A. ( 0.33, N.A.)	5	0	N.E.	0.3000 ( N.E., N.E.)	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 21 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Musculoskeletal and Connective Tissue Disorders. PT: Back Pain

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT <=450 METERS	51	15 ( 29.4)	N.A.	31	4 ( 12.9)	N.A.	2.31 ( 0.76, 6.96)	0.5857
>450 METERS	45	12 ( 26.7)	28.91 (28.91, N.A.)	18	1 ( 5.6)	N.A.	0.1275 3.79 ( 0.48, 29.74) 0.1736	
BASELINE FACIT-F FS SCORE <43 POINTS	64	20 ( 31.3)	N.A. (22.90, N.A.)	32	4 ( 12.5)	N.A.	2.32 ( 0.79, 6.83)	0.8273
>=43 POINTS	32	7 ( 21.9)	N.A. (28.91, N.A.)	16	1 ( 6.3)	N.A.	0.1152 3.09 ( 0.37, 25.75) 0.2722	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 22 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Infections and Infestations. PT: Tonsillitis

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	2 ( 2.1)	N.A.	49	6 ( 12.2)	N.A.	0.13 ( 0.03, 0.64) 0.0030	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 23 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Respiratory, Thoracic and Mediastinal Disorders. PT: Cough

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	17 ( 17.7)	N.A.	49	1 ( 2.0)	N.A.	7.58 ( 1.01, 57.03) 0.0205	
REGION								>0.9999
NORTH AMERICA & EUROPE	60	17 ( 28.3)	N.A.	30	0	N.E.	N.E.	
MIDDLE EAST	9	0	N.E.	8	0	N.E.	N.E.	
ASIA-PACIFIC	27	0	N.E.	11	1 ( 9.1)	N.A. (19.65, N.A.)	N.E.	
AGE								0.9930
<=32	35	6 ( 17.1)	N.A.	12	0	N.E.	N.E.	
>32	61	11 ( 18.0)	N.A.	37	1 ( 2.7)	N.A. (19.65, N.A.)	5.48 ( 0.71, 42.58) 0.0677	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 24 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Respiratory, Thoracic and Mediastinal Disorders. PT: Cough

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SPLENECTOMY								0.9900
YES	34	8 ( 23.5)	N.A.	26	0	N.E.	N.E.	
NO	62	9 ( 14.5)	N.A.	23	1 ( 4.3)	N.A.	2.72	
						(19.65, N.A.)	( 0.34, 21.60)	0.3236
SEX								0.9919
MALE	40	5 ( 12.5)	N.A.	23	0	N.E.	N.E.	
FEMALE	56	12 ( 21.4)	N.A.	26	1 ( 3.8)	N.A.	5.27	
							( 0.68, 40.51)	0.0742

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 25 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Respiratory, Thoracic and Mediastinal Disorders. PT: Cough

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA DIAGNOSIS								0.9909
BETA-THALASSEMIA	63	15 ( 23.8)	N.A.	34	0	N.E.	N.E.	
HEMOGLOBIN E/BETA-THALASSEMIA	28	0	N.E.	11	1 ( 9.1)	N.A. (19.65, N.A.)	N.E.	
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	11 ( 15.9)	N.A.	33	1 ( 3.0)	N.A.	4.69 ( 0.60, 36.38) 0.1036	>0.9999
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	2 ( 33.3)	N.A. ( 0.66, N.A.)	4	0	N.E.	N.E.	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	4 ( 19.0)	N.A.	12	0	N.E.	N.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 26 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Respiratory, Thoracic and Mediastinal Disorders. PT: Cough

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo		Test for Interaction P-value (4)
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)		
BASELINE NTD T/W SCORE									0.9930
<3 POINTS	30	3 ( 10.0)	N.A.	14	0	N.E.	N.E.		
>=3 POINTS	66	14 ( 21.2)	N.A.	35	1 ( 2.9)	N.A.	5.93 ( 0.78, 45.32)	0.0514	
BASELINE 6MWT									0.9929
<=450 METERS	51	10 ( 19.6)	N.A.	31	1 ( 3.2)	N.A.	5.82 ( 0.74, 45.49)	0.0570	
>450 METERS	45	7 ( 15.6)	N.A.	18	0	N.E.	N.E.		

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 27 of 40

Subgroup Analysis of Time to first Adverse Events  
 By Significant SOC/PT  
 Safety Population  
 SOC: Respiratory, Thoracic and Mediastinal Disorders. PT: Cough

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE FACIT-F FS SCORE								0.9886
<43 POINTS	64	16 ( 25.0)	N.A.	32	0	N.E.	N.E.	
>=43 POINTS	32	1 ( 3.1)	N.A.	16	1 ( 6.3)	N.A. (19.65, N.A.)	0.29 ( 0.02, 4.74) 0.3597	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 28 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Vascular Disorders. PT: Hypertension

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	19 ( 19.8)	N.A.	49	1 ( 2.0)	N.A.	9.56 ( 1.28, 71.42) 0.0070	
REGION								>0.9999
NORTH AMERICA & EUROPE	60	17 ( 28.3)	N.A.	30	1 ( 3.3)	N.A.	7.85 ( 1.04, 59.11) 0.0177	
MIDDLE EAST	9	0	N.E.	8	0	N.E.	N.E.	
ASIA-PACIFIC	27	2 ( 7.4)	N.A.	11	0	N.E.	N.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 29 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Vascular Disorders. PT: Hypertension

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								0.9942
<=32	35	2 ( 5.7)	N.A.	12	0	N.E.	N.E.	
>32	61	17 ( 27.9)	N.A.	37	1 ( 2.7)	N.A.	10.24 ( 1.36, 76.91) 0.0050	
SPLENECTOMY								0.9897
YES	34	5 ( 14.7)	N.A.	26	0	N.E.	N.E.	
NO	62	14 ( 22.6)	N.A.	23	1 ( 4.3)	N.A.	4.99 ( 0.66, 38.02) 0.0845	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 30 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Vascular Disorders. PT: Hypertension

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								0.9905
MALE	40	7 ( 17.5)	N.A.	23	0	N.E.	N.E.	
FEMALE	56	12 ( 21.4)	N.A.	26	1 ( 3.8)	N.A.	5.83 ( 0.76, 44.83) 0.0548	
BETA-THALASSEMIA DIAGNOSIS								0.9937
BETA-THALASSEMIA	63	17 ( 27.0)	N.A.	34	1 ( 2.9)	N.A.	8.83 ( 1.17, 66.40) 0.0106	
HEMOGLOBIN E/BETA-THALASSEMIA	28	2 ( 7.1)	N.A.	11	0	N.E.	N.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 31 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Vascular Disorders. PT: Hypertension

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	12 ( 17.4)	N.A.	33	0	N.E.	N.E.	>0.9999
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	0	N.E.	4	0	N.E.	N.E.	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	7 ( 33.3)	N.A. ( 8.87, N.A.)	12	1 ( 8.3)	N.A.	3.50 ( 0.43, 28.63) 0.2134	
BASELINE HEMOGLOBIN <8.5 G/DL	55	7 ( 12.7)	N.A.	29	0	N.E.	N.E.	0.9903
>=8.5 G/DL	41	12 ( 29.3)	N.A.	20	1 ( 5.0)	N.A.	6.06 ( 0.79, 46.65) 0.0482	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 32 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Vascular Disorders. PT: Hypertension

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTD-T-PRO T/W SCORE								0.9929
<3 POINTS	30	3 ( 10.0)	N.A.	14	0	N.E.	N.E.	
>=3 POINTS	66	16 ( 24.2)	N.A.	35	1 ( 2.9)	N.A.	8.22 ( 1.09, 62.03) 0.0146	
BASELINE 6MWT								0.9921
<=450 METERS	51	10 ( 19.6)	N.A.	31	0	N.E.	N.E.	
>450 METERS	45	9 ( 20.0)	N.A.	18	1 ( 5.6)	N.A.	3.12 ( 0.39, 24.75) 0.2557	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 33 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Vascular Disorders. PT: Hypertension

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE FACIT-F FS SCORE								0.9923
<43 POINTS	64	15 ( 23.4)	N.A.	32	1 ( 3.1)	N.A.	7.37 ( 0.97, 55.82)	
>=43 POINTS	32	4 ( 12.5)	N.A.	16	0	N.E.	0.0231 N.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 34 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	26 ( 27.1)	N.A.	49	5 ( 10.2)	N.A.	2.58 ( 0.99, 6.73) 0.0447	
REGION								0.9919
NORTH AMERICA & EUROPE	60	17 ( 28.3)	N.A.	30	4 ( 13.3)	N.A.	1.87 ( 0.63, 5.62) 0.2557	
MIDDLE EAST	9	5 ( 55.6)	21.55 ( 0.07, N.A.)	8	0	N.E.	N.E.	
ASIA-PACIFIC	27	4 ( 14.8)	N.A.	11	1 ( 9.1)	N.A.	1.57 ( 0.17, 14.05) 0.6856	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 35 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								0.9901
<=32	35	7 ( 20.0)	N.A.	12	0	N.E.	N.E.	
>32	61	19 ( 31.1)	N.A.	37	5 ( 13.5)	N.A.	2.25 ( 0.84, 6.04) 0.0997	
SPLENECTOMY								0.1523
YES	34	10 ( 29.4)	N.A. (19.61, N.A.)	26	1 ( 3.8)	N.A.	7.29 ( 0.93, 57.21) 0.0271	
NO	62	16 ( 25.8)	N.A.	23	4 ( 17.4)	N.A.	1.43 ( 0.48, 4.28) 0.5236	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 36 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								0.1673
MALE	40	7 ( 17.5)	N.A.	23	3 ( 13.0)	N.A.	1.16 ( 0.30, 4.58)	0.8273
FEMALE	56	19 ( 33.9)	N.A. (21.55, N.A.)	26	2 ( 7.7)	N.A.	4.70 ( 1.10, 20.21)	0.0217
BETA-THALASSEMIA DIAGNOSIS								0.5406
BETA-THALASSEMIA	63	19 ( 30.2)	N.A.	34	3 ( 8.8)	N.A.	3.42 ( 1.01, 11.59)	0.0362
HEMOGLOBIN E/BETA-THALASSEMIA	28	4 ( 14.3)	N.A.	11	1 ( 9.1)	N.A.	1.52 ( 0.17, 13.60)	0.7070

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 37 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	16 ( 23.2)	N.A.	33	2 ( 6.1)	N.A.	3.81 ( 0.87, 16.62)	0.7498
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	3 ( 50.0)	21.55 ( 3.71, N.A.)	4	1 ( 25.0)	N.A. ( 9.00, N.A.)	2.02 ( 0.21, 19.75)	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	7 ( 33.3)	N.A. ( 8.71, N.A.)	12	2 ( 16.7)	N.A. (10.18, N.A.)	0.5391 1.83 ( 0.38, 8.88)	
BASELINE HEMOGLOBIN <8.5 G/DL	55	14 ( 25.5)	N.A.	29	2 ( 6.9)	N.A.	3.30 ( 0.75, 14.59)	0.5649
>=8.5 G/DL	41	12 ( 29.3)	N.A.	20	3 ( 15.0)	N.A.	0.0946 2.05 ( 0.58, 7.28)	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 38 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTD-T-PRO T/W SCORE								0.9888
<3 POINTS	30	8 ( 26.7)	N.A.	14	0	N.E.	N.E.	
>=3 POINTS	66	18 ( 27.3)	N.A.	35	5 ( 14.3)	N.A.	1.90 ( 0.70, 5.14) 0.1977	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ibr700.sas

21JUN2022:07:09:21

Protocol: ACE-536-B-THAL-002

Page 39 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	13 ( 37.1)	N.A. ( 8.71, N.A.)	17	2 ( 11.8)	N.A. ( 9.86, N.A.)	3.16 ( 0.71, 14.04)	0.9595
>=3 TO <=5 MG/G DRY WEIGHT	23	5 ( 21.7)	N.A.	13	2 ( 15.4)	N.A. (10.18, N.A.)	1.35 ( 0.26, 7.07)	
>5 TO <=7 MG/G DRY WEIGHT	12	0	N.E.	3	0	N.E.	N.E.	
>7 TO <=15 MG/G DRY WEIGHT	14	2 ( 14.3)	N.A. (19.61, N.A.)	9	1 ( 11.1)	N.A. ( 9.00, N.A.)	0.90 ( 0.08, 10.08)	
>15 MG/G DRY WEIGHT	11	5 ( 45.5)	19.94 ( 4.07, N.A.)	5	0	N.E.	N.E.	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 40 of 40

Subgroup Analysis of Time to first Adverse Events  
By Significant SOC/PT  
Safety Population  
SOC: Skin and Subcutaneous Tissue Disorders

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT <=450 METERS	51	13 ( 25.5)	N.A.	31	3 ( 9.7)	N.A.	2.65 ( 0.76, 9.32) 0.1132	0.9386
>450 METERS	45	13 ( 28.9)	N.A. (21.55, N.A.)	18	2 ( 11.1)	N.A. (10.18, N.A.)	2.12 ( 0.47, 9.52) 0.3199	
BASELINE FACIT-F FS SCORE <43 POINTS	64	18 ( 28.1)	N.A.	32	3 ( 9.4)	N.A.	2.96 ( 0.87, 10.07) 0.0696	0.8803
>=43 POINTS	32	8 ( 25.0)	N.A.	16	1 ( 6.3)	N.A.	3.64 ( 0.45, 29.09) 0.1926	

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). MedDRA Version: 23.0 CTC Version 4.0. HR = hazard ratio; KME=Kaplan-Meier estimate.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfsocsub-ebr700.sas

21JUN2022:07:09:21

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

Protocol: ACE-536-B-THAL-002

Page 1 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	65 ( 67.7)	1.28 ( 0.13, 10.02)	49	15 ( 30.6)	N.A.	3.16 ( 1.80, 5.55) <0.0001	
REGION								0.7360
NORTH AMERICA & EUROPE	60	44 ( 73.3)	0.30 ( 0.10, 3.48)	30	10 ( 33.3)	N.A. ( 7.00, N.A.)	3.29 ( 1.65, 6.58) 0.0004	
MIDDLE EAST	9	6 ( 66.7)	10.61 ( 0.03, N.A.)	8	3 ( 37.5)	N.A. ( 1.38, N.A.)	1.99 ( 0.50, 7.98) 0.3221	
ASIA-PACIFIC	27	15 ( 55.6)	2.17 ( 0.10, N.A.)	11	2 ( 18.2)	N.A. ( 2.20, N.A.)	4.19 ( 0.96, 18.36) 0.0382	

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 2 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								
<=32	35	20 ( 57.1)	15.57 ( 1.45, N.A.)	12	3 ( 25.0)	N.A. ( 1.31, N.A.)	2.77 ( 0.82, 9.33) 0.0877	0.7283
>32	61	45 ( 73.8)	0.26 ( 0.10, 1.54)	37	12 ( 32.4)	N.A. ( 8.28, N.A.)	3.68 ( 1.94, 6.98) <0.0001	
SPLENECTOMY								
YES	34	18 ( 52.9)	14.59 ( 0.33, N.A.)	26	9 ( 34.6)	N.A. ( 4.17, N.A.)	1.85 ( 0.83, 4.13) 0.1316	0.1133
NO	62	47 ( 75.8)	0.26 ( 0.10, 2.23)	23	6 ( 26.1)	N.A.	4.67 ( 1.99, 10.94) 0.0001	

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 3 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								
MALE	40	23 ( 57.5)	10.02 ( 0.13, N.A.)	23	8 ( 34.8)	N.A. ( 2.07, N.A.)	2.02 ( 0.90, 4.54) 0.0877	0.1617
FEMALE	56	42 ( 75.0)	0.30 ( 0.10, 2.92)	26	7 ( 26.9)	N.A.	4.49 ( 2.01, 10.02) <0.0001	
BETA-THALASSEMIA DIAGNOSIS								
BETA-THALASSEMIA	63	45 ( 71.4)	1.12 ( 0.13, 10.02)	34	12 ( 35.3)	N.A. ( 4.17, N.A.)	2.84 ( 1.50, 5.39) 0.0010	0.5483
HEMOGLOBIN E/BETA-THALASSEMIA	28	16 ( 57.1)	1.81 ( 0.10, N.A.)	11	2 ( 18.2)	N.A. ( 2.20, N.A.)	4.41 ( 1.01, 19.25) 0.0300	

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 4 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	43 ( 62.3)	3.38 ( 0.26, 17.68)	33	9 ( 27.3)	N.A.	3.04 ( 1.48, 6.25) 0.0016	0.8224
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	5 ( 83.3)	0.53 ( 0.03, N.A.)	4	1 ( 25.0)	N.A. ( 8.28, N.A.)	6.58 ( 0.74, 58.72) 0.0566	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	17 ( 81.0)	0.13 ( 0.10, 1.54)	12	5 ( 41.7)	N.A. ( 0.95, N.A.)	3.65 ( 1.33, 10.00) 0.0080	
BASELINE HEMOGLOBIN <8.5 G/DL	55	35 ( 63.6)	2.92 ( 0.13, 17.68)	29	5 ( 17.2)	N.A.	4.96 ( 1.94, 12.67) 0.0002	0.1396
>=8.5 G/DL	41	30 ( 73.2)	0.26 ( 0.10, 3.48)	20	10 ( 50.0)	8.28 ( 1.38, N.A.)	2.31 ( 1.12, 4.73) 0.0224	

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 5 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTDT-PRO T/W SCORE <3 POINTS	30	18 ( 60.0)	1.12 ( 0.10, N.A.)	14	4 ( 28.6)	N.A. ( 1.38, N.A.)	3.02 ( 1.02, 8.96)	0.9923
>=3 POINTS	66	47 ( 71.2)	1.33 ( 0.13, 10.61)	35	11 ( 31.4)	N.A. ( 8.28, N.A.)	3.16 ( 1.64, 6.11)	0.0003

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 6 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON <3 MG/G DRY WEIGHT	35	24 ( 68.6)	2.23 ( 0.13, 15.57)	17	7 ( 41.2)	N.A. ( 1.31, N.A.)	2.02 ( 0.87, 4.72) 0.1005	0.8324
>=3 TO <=5 MG/G DRY WEIGHT	23	18 ( 78.3)	0.13 ( 0.10, 6.14)	13	4 ( 30.8)	N.A. ( 4.17, N.A.)	4.41 ( 1.48, 13.11) 0.0041	
>5 TO <=7 MG/G DRY WEIGHT	12	8 ( 66.7)	0.31 ( 0.03, N.A.)	3	0	N.E.	N.E.	
>7 TO <=15 MG/G DRY WEIGHT	14	8 ( 57.1)	4.48 ( 0.07, N.A.)	9	2 ( 22.2)	N.A. ( 1.31, N.A.)	3.65 ( 0.77, 17.29) 0.0847	
>15 MG/G DRY WEIGHT	11	6 ( 54.5)	17.31 ( 0.33, N.A.)	5	1 ( 20.0)	N.A. ( 8.28, N.A.)	4.00 ( 0.48, 33.54) 0.1676	

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 7 of 7

Subgroup Analysis of Time to first Bone pain or bone pain-like event  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT								
<=450 METERS	51	34 ( 66.7)	1.12 ( 0.13, 15.57)	31	8 ( 25.8)	N.A.	3.77 ( 1.74, 8.15) 0.0003	0.4437
>450 METERS	45	31 ( 68.9)	1.45 ( 0.10, 10.02)	18	7 ( 38.9)	N.A. ( 1.31, N.A.)	2.40 ( 1.05, 5.46) 0.0360	
BASELINE FACIT-F FS SCORE								
<43 POINTS	64	45 ( 70.3)	1.05 ( 0.13, 10.61)	32	11 ( 34.4)	N.A. ( 7.00, N.A.)	2.83 ( 1.46, 5.49) 0.0014	0.3932
>=43 POINTS	32	20 ( 62.5)	1.81 ( 0.10, N.A.)	16	3 ( 18.8)	N.A.	4.87 ( 1.44, 16.43) 0.0050	

Time to first bone pain or bone pain-like event is defined as the time from first dose the earliest event occurrence date. Subjects without Bone pain or bone pain-like events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Bone pain or bone pain-like events are in terms of PTs indicative of bone pain-like events, as back pain, bone pain, musculoskeletal pain, neck pain, pain in extremity, pain in jaw, and spinal pain.

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisub-ebr700.sas

21JUN2022:07:05:58

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

Protocol: ACE-536-B-THAL-002

Page 1 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
OVERALL	96	19 ( 19.8)	N.A.	49	1 ( 2.0)	N.A.	9.56 ( 1.28, 71.42) 0.0070	
REGION								>0.9999
NORTH AMERICA & EUROPE	60	17 ( 28.3)	N.A.	30	1 ( 3.3)	N.A.	7.85 ( 1.04, 59.11) 0.0177	
MIDDLE EAST	9	0	N.E.	8	0	N.E.	N.E.	
ASIA-PACIFIC	27	2 ( 7.4)	N.A.	11	0	N.E.	N.E.	

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

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

Protocol: ACE-536-B-THAL-002

Page 2 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
AGE								0.9942
<=32	35	2 ( 5.7)	N.A.	12	0	N.E.	N.E.	
>32	61	17 ( 27.9)	N.A.	37	1 ( 2.7)	N.A.	10.24 ( 1.36, 76.91) 0.0050	
SPLENECTOMY								0.9897
YES	34	5 ( 14.7)	N.A.	26	0	N.E.	N.E.	
NO	62	14 ( 22.6)	N.A.	23	1 ( 4.3)	N.A.	4.99 ( 0.66, 38.02) 0.0845	

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

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

Protocol: ACE-536-B-THAL-002

Page 3 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
SEX								0.9905
MALE	40	7 ( 17.5)	N.A.	23	0	N.E.	N.E.	
FEMALE	56	12 ( 21.4)	N.A.	26	1 ( 3.8)	N.A.	5.83 ( 0.76, 44.83)	0.0548
BETA-THALASSEMIA DIAGNOSIS								0.9937
BETA-THALASSEMIA	63	17 ( 27.0)	N.A.	34	1 ( 2.9)	N.A.	8.83 ( 1.17, 66.40)	
HEMOGLOBIN E/BETA-THALASSEMIA	28	2 ( 7.1)	N.A.	11	0	N.E.	0.0106 N.E.	

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

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

Protocol: ACE-536-B-THAL-002

Page 4 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BETA-THALASSEMIA GENE MUTATION [B0/B0, B+/B+, B+/B0] WITHOUT ALPHA-THALASSEMIA	69	12 ( 17.4)	N.A.	33	0	N.E.	N.E.	>0.9999
[B0/B0, B+/B+, B+/B0] WITH ALPHA-THALASSEMIA	6	0	N.E.	4	0	N.E.	N.E.	
[B0/B, B+/B] WITH ALPHA GENE DUPLICATION	21	7 ( 33.3)	N.A. ( 8.87, N.A.)	12	1 ( 8.3)	N.A.	3.50 ( 0.43, 28.63) 0.2134	
BASELINE HEMOGLOBIN <8.5 G/DL	55	7 ( 12.7)	N.A.	29	0	N.E.	N.E.	0.9903
>=8.5 G/DL	41	12 ( 29.3)	N.A.	20	1 ( 5.0)	N.A.	6.06 ( 0.79, 46.65) 0.0482	

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

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

Protocol: ACE-536-B-THAL-002

Page 5 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE NTDI-PRO T/W SCORE								0.9929
<3 POINTS	30	3 ( 10.0)	N.A.	14	0	N.E.	N.E.	
>=3 POINTS	66	16 ( 24.2)	N.A.	35	1 ( 2.9)	N.A.	8.22 ( 1.09, 62.03) 0.0146	

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

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

Protocol: ACE-536-B-THAL-002

Page 6 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE LIVER IRON								
<3 MG/G DRY WEIGHT	35	7 ( 20.0)	N.A.	17	1 ( 5.9)	N.A.	N.M.E.	N.M.E.
>=3 TO <=5 MG/G DRY WEIGHT	23	5 ( 21.7)	N.A.	13	0	N.E.	N.M.E.	
>5 TO <=7 MG/G DRY WEIGHT	12	3 ( 25.0)	N.A.	3	0	N.E.	N.M.E.	
>7 TO <=15 MG/G DRY WEIGHT	14	2 ( 14.3)	N.A.	9	0	N.E.	N.M.E.	
>15 MG/G DRY WEIGHT	11	2 ( 18.2)	N.A.	5	0	N.E.	N.M.E.	
			( 8.67, N.A.)					
			( 4.90, N.A.)					

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date. Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

- (1) KME of median time to first AE.
- (2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test
- (4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

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

Protocol: ACE-536-B-THAL-002

Page 7 of 7

Subgroup Analysis of Time to first Hypertension  
Safety Population

Subgroup	Luspatercept			Placebo			Luspatercept vs. Placebo	
	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	N	Patients with Event n (%)	KME [95%CI] (mon) (1)	HR [95%CI] P-value (2) (3)	Test for Interaction P-value (4)
BASELINE 6MWT								0.9921
<=450 METERS	51	10 ( 19.6)	N.A.	31	0	N.E.	N.E.	
>450 METERS	45	9 ( 20.0)	N.A.	18	1 ( 5.6)	N.A.	3.12 ( 0.39, 24.75) 0.2557	
BASELINE FACIT-F FS SCORE								0.9923
<43 POINTS	64	15 ( 23.4)	N.A.	32	1 ( 3.1)	N.A.	7.37 ( 0.97, 55.82) 0.0231	
>=43 POINTS	32	4 ( 12.5)	N.A.	16	0	N.E.	N.E.	

Time to first hypertension event is defined as the time from first dose to the earliest event occurrence date.

Subjects without hypertension events will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Hypertension events are defined in terms of PT=Hypertension

(1) KME of median time to first AE.

(2) Unstratified Cox proportional hazard model. Hazard Ratio is Luspatercept over Placebo. (3) Unstratified Log-rank Test

(4) Unstratified Cox proportional hazard model with treatment, subgroup and treatment\*subgroup interaction is to assess the significance of the interaction between treatment and the subgroup.

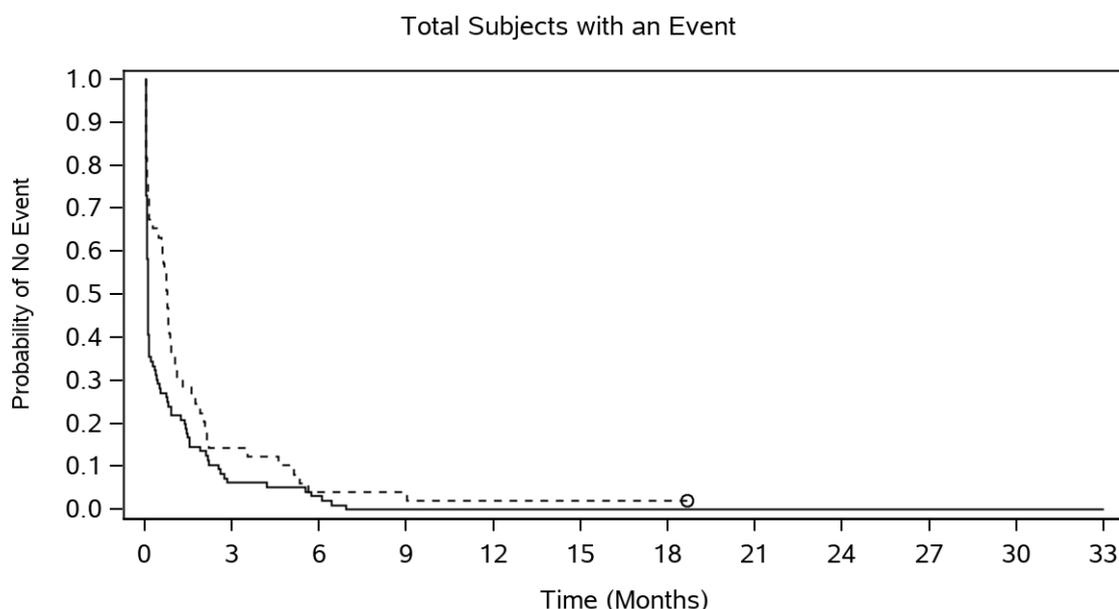
Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/tables/rt-ae-ttfaesisubhyae-ebr700.sas

04AUG2022:03:50:27

Protocol: ACE-536-B-THAL-002

Page 1 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 6 3 0 0 0 0 0 0 0 0 0

Placebo

49 7 2 2 1 1 1 0 0 0 0 0

—▲— Luspatercept (events: 96/96), median and 95% CI: 0.10 (0.07, 0.13)

--○-- Placebo (events: 48/49), median and 95% CI: 0.76 (0.46, 0.89)

Luspatercept vs. Placebo - hazard ratio: 1.47 (1.03, 2.12), pvalue: 0.0433

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTD-T-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

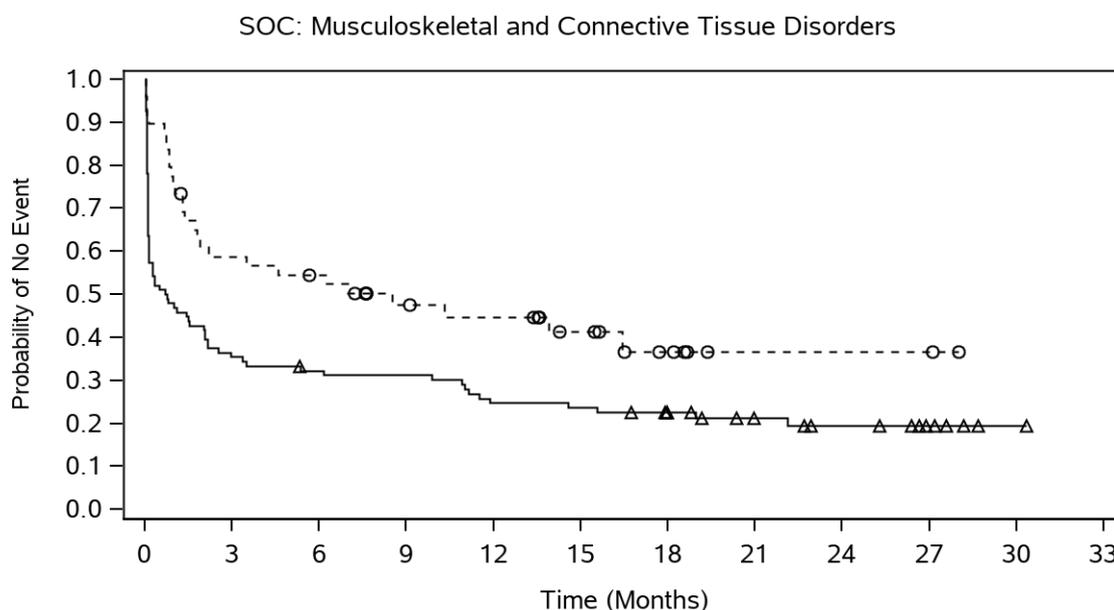
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 2 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 34 30 29 23 22 17 12 9 5 1 0

Placebo

49 28 25 18 16 11 6 2 2 2 0 0

—▲— Luspatercept (events: 76/96), median and 95% CI: 0.72 (0.13, 2.07)

--○-- Placebo (events: 28/49), median and 95% CI: 8.51 (1.74, N.A.)

Luspatercept vs. Placebo - hazard ratio: 1.85 (1.19, 2.86), pvalue: 0.0065

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

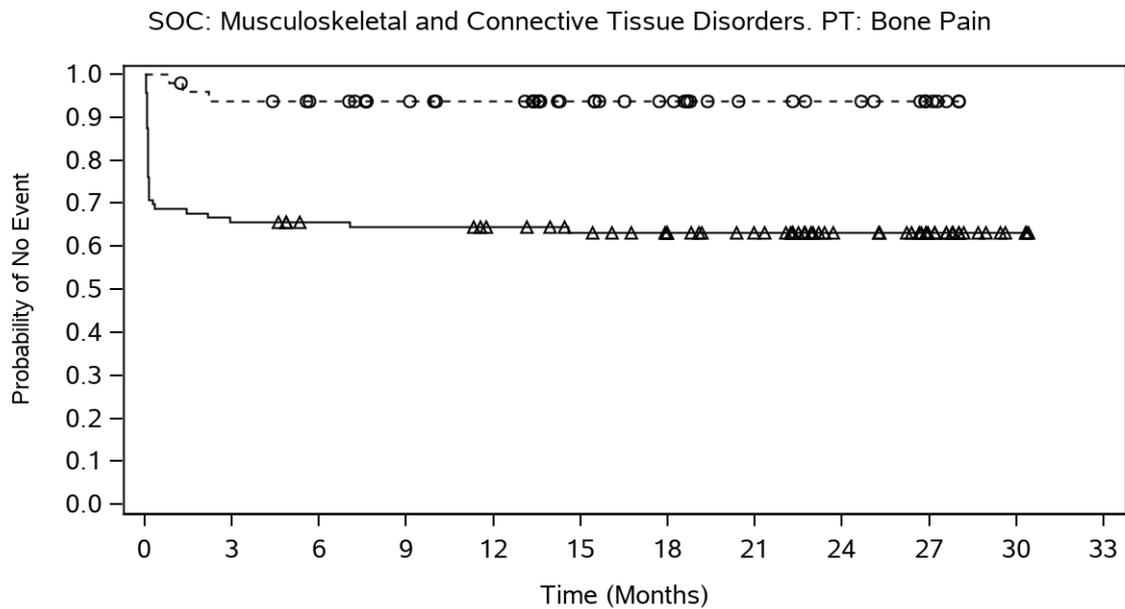
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 3 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 63 59 58 55 51 44 39 25 16 3 0

Placebo

49 45 42 36 33 24 19 12 10 5 0 0

—▲— Luspatercept (events: 35/96), median and 95% CI: N.A.

--○-- Placebo (events: 3/49), median and 95% CI: N.A.

Luspatercept vs. Placebo - hazard ratio: 7.11 (2.18, 23.15), pvalue: 0.0001

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

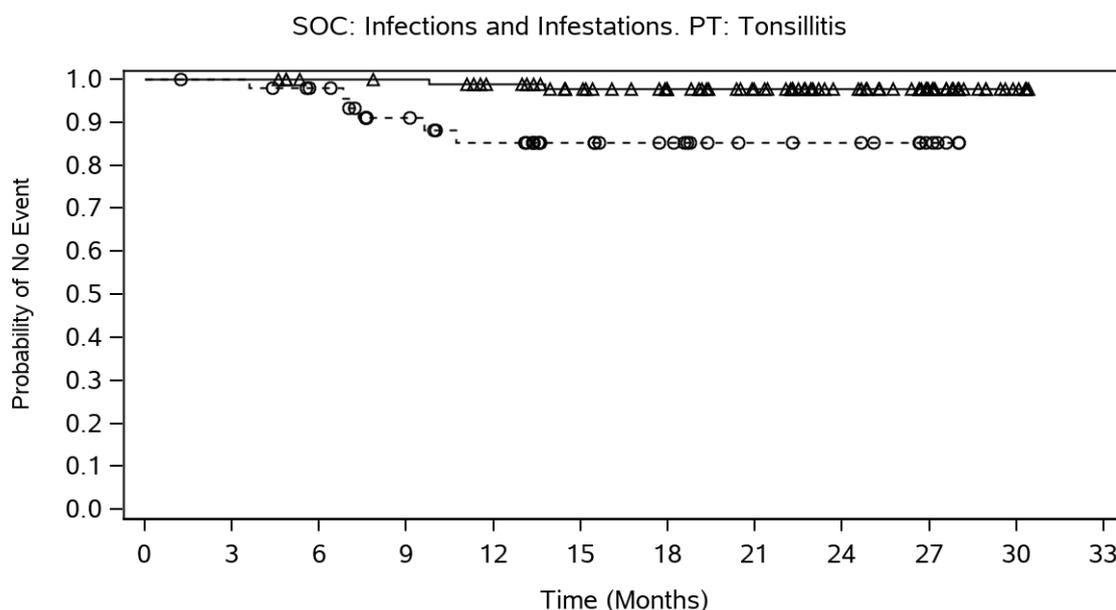
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 4 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 96 92 91 86 78 68 59 43 27 5 0

Placebo

49 48 44 34 29 21 17 11 10 5 0 0

—▲— Luspatercept (events: 2/96), median and 95% CI: N.A.

--○-- Placebo (events: 6/49), median and 95% CI: N.A.

Luspatercept vs. Placebo - hazard ratio: 0.13 (0.03, 0.64), pvalue: 0.0033

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

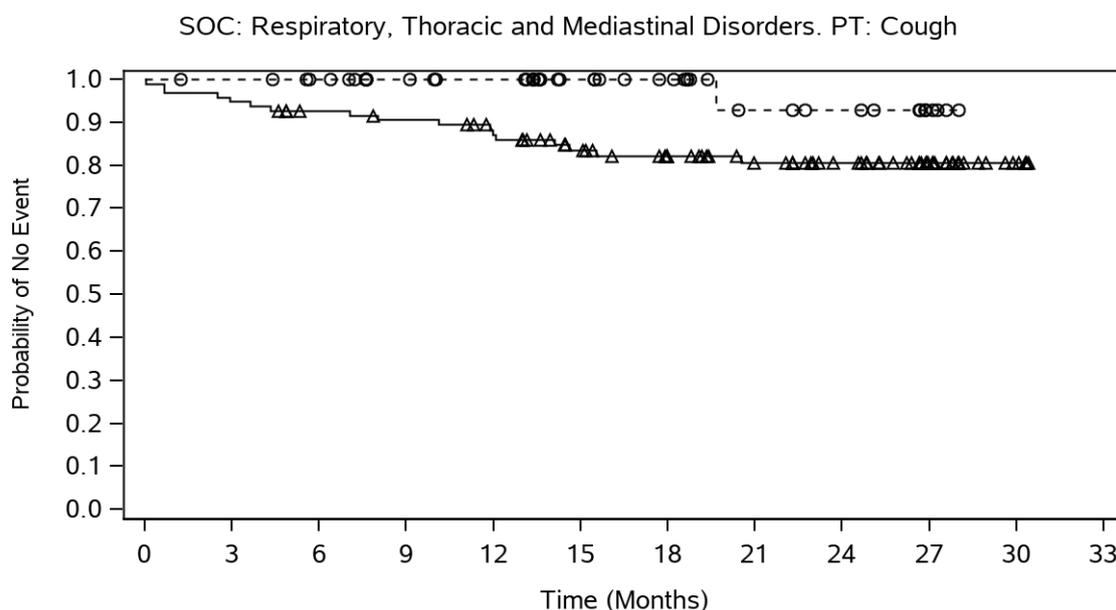
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 5 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 91 85 82 76 66 56 48 39 23 5 0

Placebo

49 48 45 38 35 25 20 12 10 4 0 0

—▲— Luspatercept (events: 17/96), median and 95% CI: N.A.

-○- Placebo (events: 1/49), median and 95% CI: N.A.

Luspatercept vs. Placebo - hazard ratio: 7.24 (0.96, 54.68), pvalue: 0.0248

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

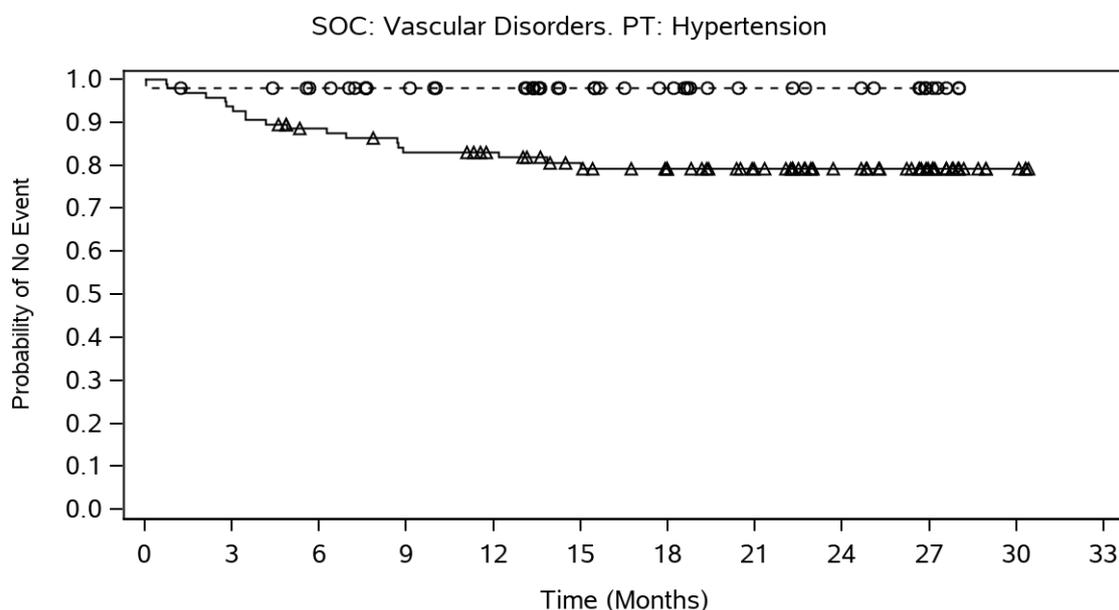
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 6 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96	90	81	75	71	64	56	48	35	22	4	0
----	----	----	----	----	----	----	----	----	----	---	---

Placebo

49	47	44	38	35	25	20	13	11	5	0	0
----	----	----	----	----	----	----	----	----	---	---	---

—▲— Luspatercept (events: 19/96), median and 95% CI: N.A.

-○- - Placebo (events: 1/49), median and 95% CI: N.A.

Luspatercept vs. Placebo - hazard ratio: 9.32 (1.25, 69.66), pvalue: 0.0079

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

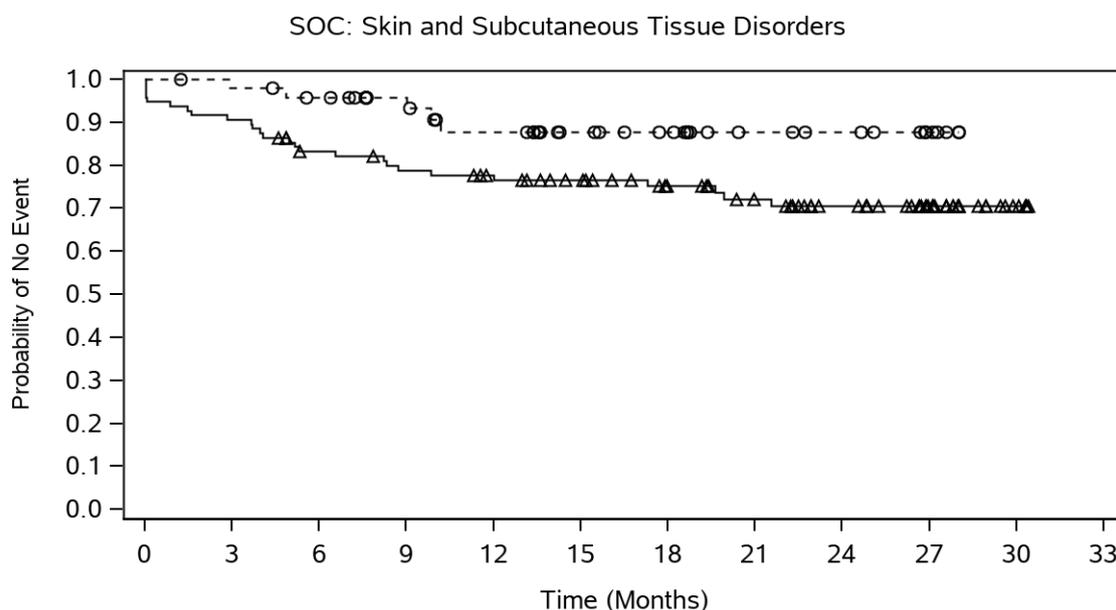
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 7 of 7

Figure 05.1:  
Kaplan-Meier Plot of Time to First Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96	87	76	71	67	61	52	45	36	24	5	0
----	----	----	----	----	----	----	----	----	----	---	---

Placebo

49	47	44	37	31	23	19	12	10	5	0	0
----	----	----	----	----	----	----	----	----	---	---	---

—▲— Luspatercept (events: 26/96), median and 95% CI: N.A.

--○-- Placebo (events: 5/49), median and 95% CI: N.A.

Luspatercept vs. Placebo - hazard ratio: 2.59 (0.99, 6.77), pvalue: 0.0434

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

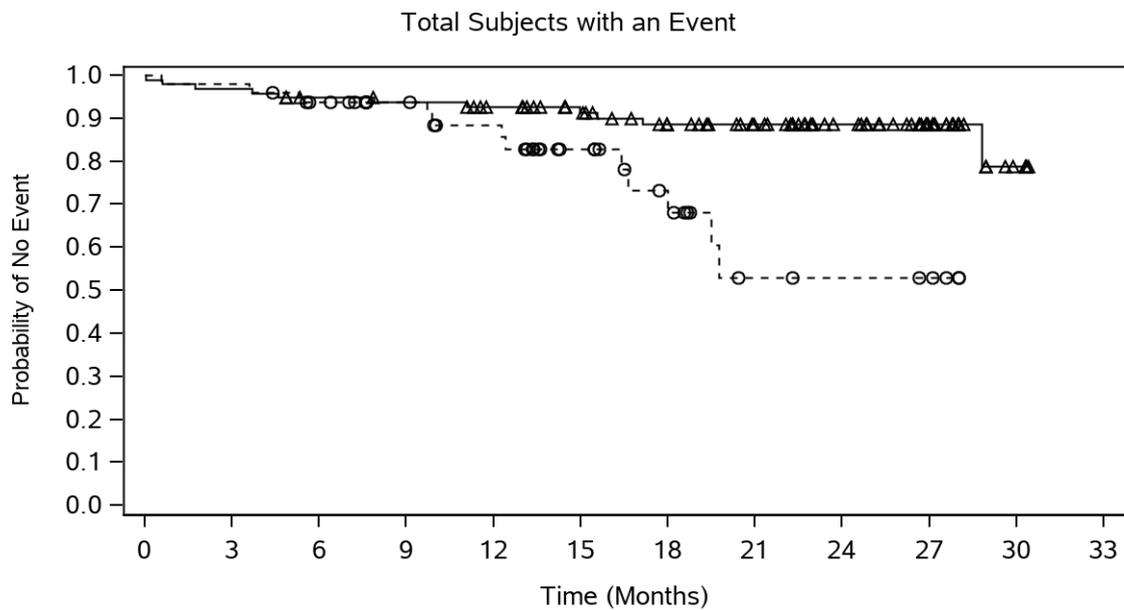
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:32:14

Protocol: ACE-536-B-THAL-002

Page 1 of 2

Figure 05.3:  
Kaplan-Meier Plot of Time to First Serious Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 93 88 86 81 73 62 54 39 23 4 0

Placebo

49 48 43 36 31 21 14 6 5 4 0 0

—▲— Luspatercept (events: 11/96), median and 95% CI: N.A.

-○- - Placebo (events: 12/49), median and 95% CI: N.A. (18.00, N.A.)

Luspatercept vs. Placebo - hazard ratio: 0.29 (0.12, 0.69), pvalue: 0.0032

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

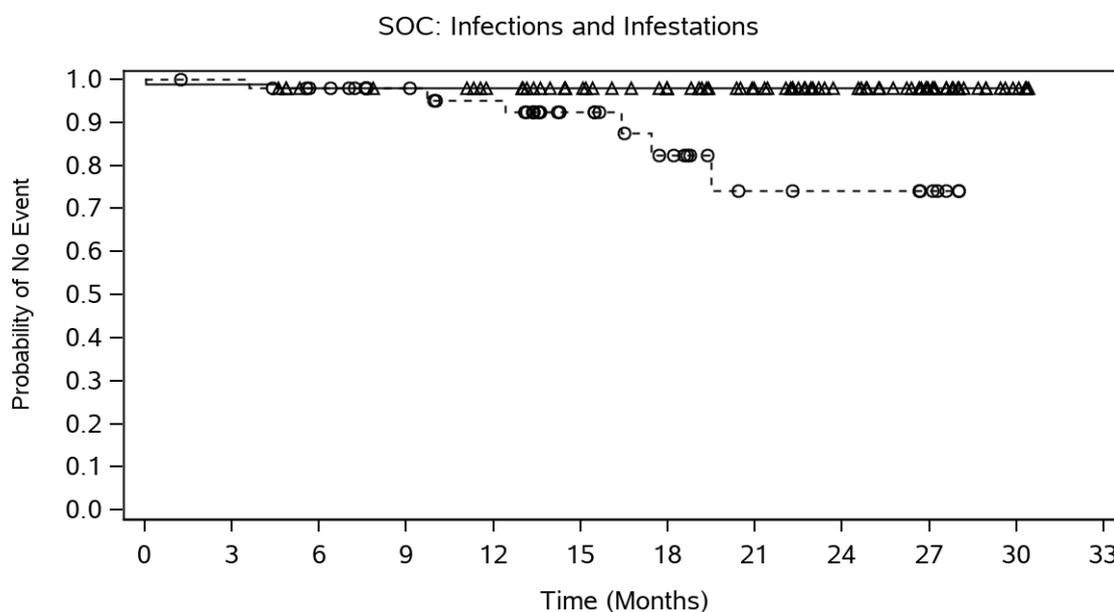
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:33:34

Protocol: ACE-536-B-THAL-002

Page 2 of 2

Figure 05.3:  
Kaplan-Meier Plot of Time to First Serious Adverse Events - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 95 90 89 85 77 68 59 43 27 5 0

Placebo

49 48 44 37 33 22 15 8 7 5 0 0

—▲— Luspatercept (events: 2/96), median and 95% CI: N.A.

--○-- Placebo (events: 6/49), median and 95% CI: N.A. (19.48, N.A.)

Luspatercept vs. Placebo - hazard ratio: 0.11 (0.02, 0.57), pvalue: 0.0017

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

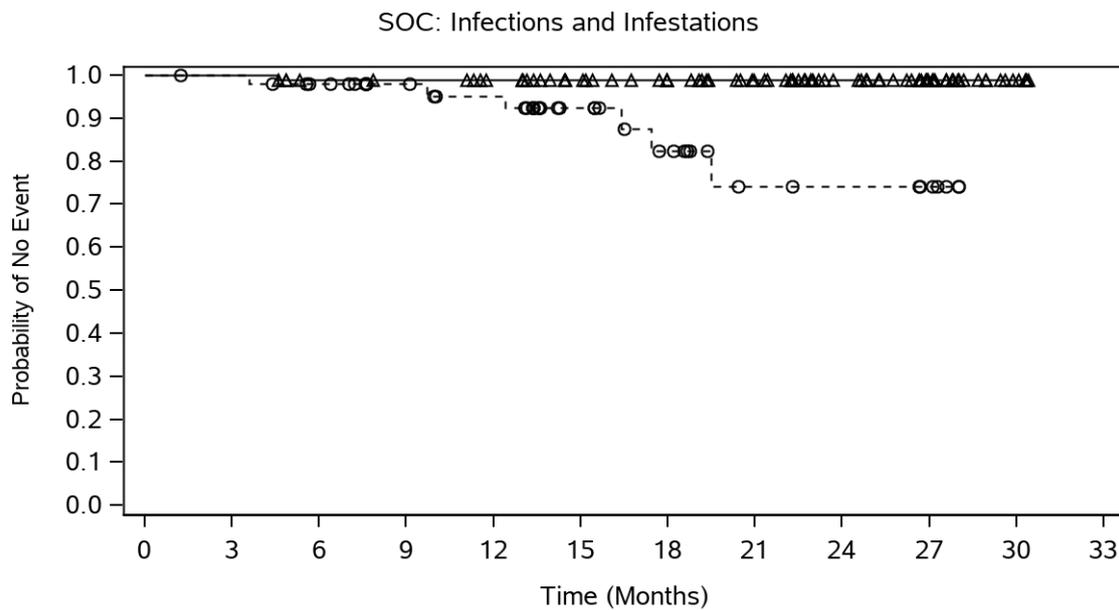
Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:33:34

Protocol: ACE-536-B-THAL-002

Page 1 of 1

Figure 05.2:  
Kaplan-Meier Plot of Time to First Adverse Events CTC Grade  $\geq 3$  - By Significant SOC/PT - Safety Population



Number of Subjects at Risk

Luspatercept

96 96 91 90 86 78 69 60 44 27 5 0

Placebo

49 48 44 37 33 22 15 8 7 5 0 0

—▲— Luspatercept (events: 1/96), median and 95% CI: N.A.

- -○- - Placebo (events: 6/49), median and 95% CI: N.A. (19.48, N.A.)

Luspatercept vs. Placebo - hazard ratio: 0.05 (0.01, 0.42), pvalue: 0.0002

The time to first AE is defined as the time from first dose to the first occurrence of the AE of the relevant type. Subjects without a relevant AE will be censored 63 days after the last dose, at study discontinuation, or at data cut-off (whichever occurred first). Cox proportional hazards ratio and log-rank test stratified by baseline Hb category ( $\geq 8.5$  g/dL,  $< 8.5$ g/dL) and baseline NTDT-PRO T/W score category ( $\geq 3$  vs.  $< 3$ ).

Program Source: /opt/zfs001/prd/bms214671/stats/market/prog/figures

Program Name: rg-ae-ttfsocae-ebr700.sas

21JUN2022:10:33:03