

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

*Cerliponase alfa (Brineura<sup>®</sup>)*

BioMarin International Ltd.

**Modul 4 A**

*Neuronale Ceroid-Lipofuszinose (NCL) Typ 2, auch als  
Tripeptidylpeptidase 1 (TPP1)-Mangel bezeichnet*

Medizinischer Nutzen und  
medizinischer Zusatznutzen,  
Patientengruppen mit therapeutisch  
bedeutsamem Zusatznutzen

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>4</b>
<b>Abbildungsverzeichnis</b> .....	<b>8</b>
<b>Abkürzungsverzeichnis</b> .....	<b>10</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 .....	27
4.2.3 Informationsbeschaffung .....	28
4.2.3.1 Studien des pharmazeutischen Unternehmers .....	28
4.2.3.2 Bibliografische Literaturrecherche .....	28
4.2.3.3 Suche in Studienregistern .....	29
4.2.3.4 Selektion relevanter Studien .....	30
4.2.4 Bewertung der Aussagekraft der Nachweise .....	30
4.2.5 Informationssynthese und -analyse .....	31
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien .....	31
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien .....	32
4.2.5.3 Meta-Analysen .....	32
4.2.5.4 Sensitivitätsanalysen .....	33
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren .....	34
4.2.5.6 Indirekte Vergleiche .....	35
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen .....	37
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel .....	37
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel .....	37
4.3.1.1.1 Studien des pharmazeutischen Unternehmers .....	37
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche .....	38
4.3.1.1.3 Studien aus der Suche in Studienregistern .....	40
4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel .....	41
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel .....	42
4.3.1.2.1 Studiendesign und Studienpopulationen .....	42
4.3.1.2.2 Verzerrungspotenzial auf Studienebene .....	44
4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien .....	45
4.3.1.3.1 <Endpunkt xxx> – RCT .....	45
4.3.1.3.2 Subgruppenanalysen – RCT .....	48
4.3.1.3.3 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien .....	48
4.3.2 Weitere Unterlagen .....	49
4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien .....	49

4.3.2.1.1	Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche .....	49
4.3.2.1.2	Charakteristika der Studien für indirekte Vergleiche.....	49
4.3.2.1.3	Ergebnisse aus indirekten Vergleichen .....	50
4.3.2.1.3.1	<Endpunkt xxx> – indirekte Vergleiche aus RCT .....	50
4.3.2.1.3.2	Subgruppenanalysen – indirekte Vergleiche aus RCT .....	52
4.3.2.2	Nicht randomisierte vergleichende Studien .....	53
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien .....	53
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	53
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien .....	54
4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien.....	54
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien .....	55
4.3.2.3	Weitere Untersuchungen.....	55
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen .....	56
4.3.2.3.2	Charakteristika der weiteren Untersuchungen .....	56
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen .....	56
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen .....	56
4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen .....	57
4.3.2.4	Zusammenfassung der Ergebnisse aus weiteren Unterlagen .....	57
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	57
4.4.1	Beurteilung der Aussagekraft der Nachweise .....	57
4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	58
4.4.2.1	Methoden .....	58
4.4.2.2	Charakteristika der eingeschlossenen Studien .....	92
4.4.2.3	Ergebnisse aus weiteren Untersuchungen.....	121
4.4.2.3.1	Krankheitsprogression auf CLN2-spezifischen Bewertungsskalen .....	122
4.4.2.3.1.1	Mittlere Änderung der ML/HML-Skala von Baseline in den Studien 190-201/202 und 190-901 (historische Kontrolle); deskriptive Analyse .....	130
4.4.2.3.1.2	Patientenrelevante Änderung der ML/HML-Bewertungsskala .....	136
4.4.2.3.1.3	Primäre Analyse: Abfall des CLN2-Scores um zwei Punkte (Responder- und Ereigniszeitanalysen) .....	136
4.4.2.3.1.4	Progressionsrate (Verschlechterung des ML/HML-Skalenwerts) .....	145
4.4.2.3.1.5	Einzeldomänen der ML/HML-Skala und andere Summenscores .....	147
4.4.2.3.1.6	Ergebnisse der weiteren Nachverfolgung .....	150
4.4.2.3.2	Lebensqualität gemessen mittels PedsQL™ und CLN2 spezifischem Fragebogen - weitere Untersuchungen.....	154
4.4.2.3.3	Verträglichkeit und Sicherheit (UEs) – weitere Untersuchungen .....	159
4.4.2.3.4	Subgruppenanalysen – weitere Untersuchungen.....	185
4.4.2.4	Zusammenfassung der Ergebnisse aus weiteren Unterlagen .....	186
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht .....	204
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte .....	204
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	204

4.5.2	Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen.....	204
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen .....	205
4.5.4	Verwendung von Surrogatendpunkten .....	205
4.6	Liste der eingeschlossenen Studien.....	206
4.7	Referenzliste.....	207
<b>Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche .....</b>		<b>214</b>
<b>Anhang 4-B : Suchstrategien – Suche in Studienregistern.....</b>		<b>216</b>
<b>Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche).....</b>		<b>217</b>
<b>Anhang 4-D : Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern).....</b>		<b>218</b>
<b>Anhang 4-E : Methodik der eingeschlossenen Studien.....</b>		<b>219</b>
<b>Anhang 4-F : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten .....</b>		<b>254</b>

**Tabellenverzeichnis**

	<b>Seite</b>
Tabelle 4-1: Matrix der patientenrelevanten Endpunkte der eingeschlossenen Studie 190-201/202 mit dem zu bewertenden Arzneimittel .....	18
Tabelle 4-2: Zusammenfassung der Ergebnisse für Krankheitsprogression gemessen mittels ML/HML-Skala: Cerliponase alfa vs. historische Kontrolle nach 1:1 Paarbildung entsprechend klinisch relevanter prognostischer Faktoren .....	20
Tabelle 4-3: Überblick der Ergebnisse zur Lebensqualität (PedsQL™; ITT, n=23) .....	23
Tabelle 4-4: Überblick der Ergebnisse zur Lebensqualität (CLN2-spezifisches QoL-Instrument; ITT, n=23).....	23
Tabelle 4-5: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel .....	38
Tabelle 4-6: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	38
Tabelle 4-7: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel .....	41
Tabelle 4-8: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	42
Tabelle 4-9: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel .....	43
Tabelle 4-10: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	44
Tabelle 4-11: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel.....	44
Tabelle 4-12: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	45
Tabelle 4-13: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel .....	45
Tabelle 4-14: Operationalisierung von <Endpunkt xxx>.....	46
Tabelle 4-15: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT mit dem zu bewertenden Arzneimittel .....	47
Tabelle 4-16: Ergebnisse für <Endpunkt xxx> aus RCT mit dem zu bewertenden Arzneimittel.....	47
Tabelle 4-17: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche .....	50
Tabelle 4-18: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	50
Tabelle 4-19: Operationalisierung von <Endpunkt xxx>.....	51
Tabelle 4-20: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche .....	51

Tabelle 4-21: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche.....	52
Tabelle 4-22: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien .....	54
Tabelle 4-23: Operationalisierung von <Endpunkt xxx>.....	54
Tabelle 4-24: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien .....	55
Tabelle 4-25: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen.....	56
Tabelle 4-26: Liste der Studien des pharmazeutischen Unternehmers – Weitere Untersuchungen.....	62
Tabelle 4-27: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – Weitere Untersuchungen .....	62
Tabelle 4-28: Hamburg und WCMC-Bewertungsskalen zur quantitativen Erfassung der Krankheitsaktivität .....	63
Tabelle 4-29: Vergleich der HML Skala mit der ML Skala .....	65
Tabelle 4-30: Leitsymptome der Ceroid-Lipofuszinose Typ 2 -zeitlicher Verlauf .....	78
Tabelle 4-31: Auswertbare Patienten des DEM-CHILD Registers .....	89
Tabelle 4-32: Abfall (Rate) des HML-Scores bei unbehandelten CLN2-Patienten (Punkte/48 Wochen; auswertbare Patienten) [56].....	90
Tabelle 4-33: MMRM-Analysen des Abfalls (Rate) des HML-Scores bei unbehandelten CLN2-Patienten (Punkte/48 Wochen; auswertbare Patienten) [56] .....	91
Tabelle 4-34: Charakterisierung der eingeschlossenen Studien.....	94
Tabelle 4-35: Charakterisierung der Interventionen – Weitere Studien mit dem zu bewertenden Arzneimittel [37].....	97
Tabelle 4-36: Charakterisierung der Studienpopulationen – Weitere Untersuchungen mit dem zu bewertenden Arzneimittel und Untersuchungen zum natürlichen Verlauf .....	99
Tabelle 4-37: Baseline-Werte der auswertbaren Patientenpopulation (190-901) auf der HML-, HMLV- und HMLVS-Skala vor und nach 1:1 Paarbildung mit Patienten der Studie 190-201 [57].....	105
Tabelle 4-38: Matrix der Endpunkte in den eingeschlossenen Studien mit dem zu bewertenden Arzneimittel .....	122
Tabelle 4-39: Operationalisierung des Endpunkts „Krankheitsprogression auf CLN2-spezifischen Bewertungsskalen.....	122
Tabelle 4-40: Bewertung des Verzerrungspotenzials des Endpunkts „Krankheitsprogression auf CLN2 spezifischen Bewertungsskalen (ML/HML-Skala“ mit dem zu bewertenden Arzneimittel – weitere Untersuchungen.....	127
Tabelle 4-41: Geschätzte Abnahme des ML/HML Wertes über 48 Wochen: Cerliponase alfa (Studie 190-201) vs. unbehandelte Kontrolle (Studie 190-901), mit oder ohne Paarbildung.....	132
Tabelle 4-42: Primäre Analyse der Ansprechrate gemessen mittels ML/HML-Skala (<2 oder ≥2 Punkte/48 Wochen): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) nach 1:1 Paarbildung [57] .....	137

Tabelle 4-43: Ansprechrate gemessen mittels ML/HML-Skala (<2 oder $\geq 2$ Punkte/48 Wochen) über 1.000 Simulationen zur 1:1 Paarbildung („matching“): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten; Sensitivitätsanalyse) [57] .....	138
Tabelle 4-44: Sensitivitätsanalysen der primären Responder-Analyse (<2 oder $\geq 2$ Punkte / 48 Wochen) [57].....	139
Tabelle 4-45: Ansprechrate getrennt nach motorischer und Sprachdomäne auf der ML/HML-Skala (<1 oder $\geq 1$ Punkte/48 Wochen): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) nach 1:1 Paarbildung [57] .....	141
Tabelle 4-46: Kaplan-Meier-Analysen der Krankheitsprogression gemessen mittels ML/HML-Skala: Cerliponase alfa (190-201) vs. unbehandelte Kontrolle (190-901) nach 1:1 Paarbildung [57].....	142
Tabelle 4-47: Sensitivitätsanalysen der primären Kaplan-Meier Analyse mit 1:1 Paarbildung [57].....	144
Tabelle 4-48: Geschätzte Abnahme des ML/HML-Scores/48 Wochen: Cerliponase alfa (190-201) vs. unbehandelte Kontrolle (190-901) mit und ohne 1:1 Paarbildung [57] .....	146
Tabelle 4-49: Geschätzte Abnahme der motorischen bzw. Sprachdomäne (jeweils 0-3 Punkte)/48 Wochen: Cerliponase alfa (190-201, ITT) vs. unbehandelte Kontrolle (190-901) mit 1:1 Paarbildung [57] .....	148
Tabelle 4-50: Geschätzte Abnahme des MLV/HMLV-Scores (0-9 Punkte)/48 Wochen: Cerliponase alfa (190-201) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) ohne Paarbildung [57] .....	149
Tabelle 4-51: Geschätzte Abnahme des Gesamtscores (0-12 Punkte)/48 Wochen: Cerliponase alfa (190-201/202 ITT, 300 mg-Dosisintervall) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) ohne Paarbildung [57] .....	150
Tabelle 4-52: Ansprechrate gemessen mittels ML/HML-Skala (<2 oder $\geq 2$ Punkte/48 Wochen): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) nach 1:1 Paarbildung; DCO: 01 Nov 2016 [59] .....	151
Tabelle 4-53: Operationalisierung des Endpunkts gesundheitsbezogene Lebensqualität (PedsQL™) – weitere Untersuchungen .....	154
Tabelle 4-54: Änderung des PedsQL™ von Baseline bis zur letzten Erhebung (ITT n=23, gesamtes Dosisintervall) [26,57].....	157
Tabelle 4-55: Operationalisierung des Endpunkts „CLN2-spezifische Lebensqualität“ – weitere Untersuchungen .....	158
Tabelle 4-56: Änderung der gesundheitsbezogenen Lebensqualität von Baseline bis zur letzten Erhebung gemessen mittels eines CLN2-spezifischen Instruments (gesamtes Dosisintervall) [26,57] .....	159
Tabelle 4-57: Operationalisierung des Endpunkts Verträglichkeit und Sicherheit.....	159
Tabelle 4-58: Überblick der in der Studie 190-201/202 aufgetretenen, unerwünschten Ereignisse in den Kohorten 1-3, während der stabilen Dosierung und in der Gesamtpopulation [26].....	166

Tabelle 4-59: Anzahl der Patienten, die mindestens ein unerwünschtes Ereignis entwickelten („Sicherheits“-Population) [26,34] .....	166
Tabelle 4-60: Inzidenz unerwünschter Ereignisse stratifiziert nach Systemorganklassen („Sicherheits“-Population) [26,34].....	167
Tabelle 4-61: Inzidenz sehr häufiger unerwünschter Ereignisse ( $\geq 10\%$ der Patienten) mit Angabe der Inzidenz schwerer UEs (CTCAE Grad $\geq 3$ ; „Sicherheits“-Population) [26,34] .	169
Tabelle 4-62: Inzidenz schwerwiegender UEs und Todesfälle („Sicherheits“-Population) [26,27,34] .....	171
Tabelle 4-63: Inzidenz der CTCAE Grad $\geq 3$ UEs pro SOK („Sicherheits“-Population) [26,34] .....	172
Tabelle 4-64: Studienabbrüche aufgrund eines UE („Sicherheits“-Population) [26,34] .....	173
Tabelle 4-65: Präspezifizierte UEs besonderen Interesses: Inzidenz von Überempfindlichkeitsreaktionen („Sicherheits“-Population) [26,34].....	175
Tabelle 4-66: Präspezifizierte UEs besonderen Interesses: Inzidenz von UEs, die bei mindestens drei Patienten (auf SOK- und PT-Ebene) in einem zeitlichen Zusammenhang mit der Infusion von Cerliponase alfa auftraten („Sicherheits“-Population) [26,34] .....	177
Tabelle 4-67: UEs besonderen Interesses: Inzidenz von UEs innerhalb der SOK „Erkrankungen des Nervensystems“ und Konvulsionen („Sicherheits“-Population, gesamtes Dosisintervall) [26,34] .....	179
Tabelle 4-68: UEs, die mit dem ICV-Zugang in Zusammenhang stehen („Sicherheits“-Population, gesamtes Dosisintervall) [26,34].....	182
Tabelle 4-69: Geschätzte Rate des Abfalls des ML-Scores (Punkte pro 48 Wochen) nach Studienzentrum („Wirksamkeits“-Population, 300 mg-Dosisintervall) [26] .....	185
Tabelle 4-70: Zusammenfassung der Ergebnisse patientenrelevanter Wirksamkeits-Endpunkte.....	186
Tabelle 4-71: Überblick der Ergebnisse zur Lebensqualität (PedsQL™; ITT, n=23) .....	189
Tabelle 4-72: Überblick der Ergebnisse zur Lebensqualität (CLN2-spezifisches QoL-Instrument; ITT, n=23).....	190
Tabelle 4-73: Überblick der in der Studie 190-201/202 aufgetretenen unerwünschten Ereignisse (behandlungsspezifische UEs, n=24, „Sicherheits“-Population) .....	191
Tabelle 4-74: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	204
Tabelle 4-75 (Anhang): Studiendesign und -methodik für Studie 190-201 .....	220
Tabelle 4-76 (Anhang): Studiendesign und -methodik für Studie 190-202 (fortlaufende Studie) [45].....	240
Tabelle 4-77 (Anhang): Studiendesign und -methodik der Studie 190-901 (190-901 Supplemental Report zum natürlichen Verlauf der CLN2) und des historischen Vergleichs mit den Studien 190-201/202 [56] .....	247
Tabelle 4-78 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie 190-201 .....	255



**Abbildungsverzeichnis**

	<b>Seite</b>
Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel .....	40
Abbildung 2: Meta-Analyse für <Endpunkt xxx> aus RCT; <zu bewertendes Arzneimittel> versus <Vergleichstherapie> .....	48
Abbildung 3: Typischer Verlauf der Ceroid-Lipofuszinose Typ 2 .....	80
Abbildung 4: HML-Score nach Alter bei Patienten der Hamburg Kohorte (n=29).....	83
Abbildung 5: Querschnittsanalyse der motorischen (links) und Sprachdomäne (rechts) der HML-Skala, Hamburg Kohorte (n=29).....	84
Abbildung 6: Querschnittsanalyse der Domäne „Sehvermögen“ der HML-Skala, Hamburg Kohorte (n=29).....	85
Abbildung 7: Querschnittsanalyse der Domäne “Krampfanfälle” der HML-Skala, Hamburg Kohorte (n=29).....	86
Abbildung 8: Summenscore der kombinierten motorischen Skala und Sprachskala in unabhängigen Patienten-Kohorten, gemessen mittels unterschiedlicher Instrumente: WCMC-Kohorten und Hamburg Kohorte.....	87
Abbildung 9: Individueller Abfall des HML-Scores (Punkte) während der linearen Phase der Krankheitsprogression (Schätzung unter Einsatz des Algorithmus erste/letzte Erhebung; auswertbare Patienten) [56] .....	92
Abbildung 10: Studiendesign von Studie 190-201 und Dosierung von Cerliponase alfa während der Dosisescalation und der stabilen Dosierungsphase.....	109
Abbildung 11: Mittlere Änderung der ML/HML-Skala bezogen auf Baseline (Studie 190-201/202, ITT-Population; 300 mg-Dosisintervall / Studie 190-901 auswertbare Patienten). 134	134
Abbildung 12: Mittlere Änderung der MLV/HMLV-Skala bezogen auf Baseline (Studie 190-201/202, ITT-Population; 300 mg-Dosisintervall / Studie 190-901 auswertbare Patienten).....	135
Abbildung 13: Mittlere Änderung der Gesamt-Scores der MLVS/HMLVS-Skala bezogen auf Baseline (Studie 190-201/202, ITT-Population; 300 mg-Dosisintervall / Studie 190-901 auswertbare Patienten) .....	135
Abbildung 14: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Abnahme um $\geq 2$ Punkte oder Wert von Null): Cerliponase alfa (190-201/202; ITT) vs. unbehandelte Kontrolle (190-901) nach 1:1 Paarbildung (Alter, ML/HML Wert) [57].....	143
Abbildung 15: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Zeit bis zur ersten dauerhaften Abnahme um $\geq 2$ Punkte): Cerliponase alfa (190-201/202; ITT) vs. unbehandelte Kontrolle (190-901; Auswertbare Patienten, n=41) ohne Paarbildung [57].....	145
Abbildung 16: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Abnahme um $\geq 2$ Punkte oder Wert von Null): Cerliponase alfa (190-	

201/202) vs. unbehandelte Kontrolle (190-901) mit 1:1 Paarbildung (Alter, ML/HML Wert); DCO: 01 Nov 2016 [59] .....	152
Abbildung 17: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Abnahme um $\geq 2$ Punkte oder Wert von Null): Cerliponase alfa (190-201/202) vs. unbehandelte Kontrolle (190-901) mit 1:1 Paarbildung (Alter, Mutation, ML/HML Wert); DCO: 01 Nov 2016 [59] .....	153
Abbildung 18: Anzahl der Überempfindlichkeitsreaktionen pro Patient („Sicherheits“-Population) [26].....	176
Abbildung 19: Anzahl der Patienten, bei denen eine oder mehrere Konvulsionen aufgetreten waren (SMQ „Konvulsionen“, „Sicherheits“-Population) [26] .....	181
Abbildung 20: Patientenfluss der Studie 190-201 und der Folgestudie 190-202 .....	246

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
%	Prozent
AEOSI	Adverse Event of Special Interest (Unerwünschtes Ereignis von besonderem Interesse)
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BL	Baseline
BMN-190	Cerliponase alfa
CHMP	Committee for Medicinal Products for Human Use (Ausschuss für Humanarzneimittel)
CLN	Ceroid-Lipofuszinose
CLN2	Ceroid-Lipofuszinose Typ 2
CONSORT	Consolidated Standards of Reporting Trials
CSF	Cerebrospinal fluid (Liquor cerebrospinalis)
CSR	Clinical Study Report (Klinischer Studienbericht)
CTCAE	Common Terminology Criteria for Adverse Events (Allgemeine Toxizitätskriterien)
d. h.	Das heißt
DIMDI	Deutsches Institut für Medizinische Dokumentation
DMC	Data Monitoring Committee (Datenüberwachungskomitee)
EEG	Elektroenzephalogramm
EG	Europäische Gemeinschaft
EMA	European Medicines Agency (Europäische Arzneimittel-Agentur)
ERT	Enzyme Replacement Therapy (Enzymersatzthreapie)
EU	Europäische Union
FDA	Food and Drug Association
GCP	Good Clinical Practice (Gute klinische Praxis)
GRADE	Grading of Recommendations Assessment, Development and Evaluation
H <sub>0</sub>	Nullhypothese
H <sub>1</sub>	Alternativhypothese
H <sub>A</sub>	Alternativhypothese

HML	Hamburg Motor/Language (Kombinierte Skala Motorisch/Sprache, entwickelt in Hamburg)
HR	Hazard-Ratio
i.v.	Intravenös
IAR	Infusion-Associated Reaction (Reaktion in Verbindung mit einer Infusion)
ICH E9	International Conference on Harmonization, E9 (Internationale Konferenz zur Harmonisierung, E9)
ICV	Intrazerebrovaskulär
IgE	Immunglobulin E
IGWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
iRR	Inverses Relatives Risiko
ISE	Integrated Summary of Efficacy
ISPOR	International Society For Pharmacoeconomics and Outcomes Research
ITT	Intention to treat
ITT	Intention-to-treat
$k\omega$	Weighted Kappa (Gewichteter Kappa-Koeffizient)
KI	Konfidenzintervall
L	Language (Sprache)
LINCL	Late Infantile Neuronal Ceroid Lipofuscinosis (Spätinfantile, neuronale Ceroid-Lipofuszinose)
LOCB	Last Observation Carried Backward
LOCF	Last Observation Carried Forward
M	Motorisch
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities (Medizinisches Wörterbuch für Aktivitäten im Rahmen der Arzneimittelzulassung)
mg	Milligramm
MID	Minimal Important Difference (Kleinster relevanter Unterschied)
Min	Minimum
ML	Motor/Language (Kombinierte Skala Motorisch/Sprache)
ml	Milliliter
MLV	Motor/Language/Vision (Kombinierte Skala motorisch/Sprache/Sehen)

MLVS	Motor/Language/Vision/Seizures (Kombinierte Skala Motorisch/Sprache/Sehen/Krampfanfälle)
mm <sup>2</sup>	Quadratmillimeter
mm <sup>3</sup>	Kubikmillimeter
MMRM	Mixed Model for Repeated Measures (Gemischtes Modell mit Messwiederholung)
MRI	Magnetic Resonance Imaging (Magnetresonanztomographie)
MRT	Magnetresonanztomographie
MTC	Mixed Treatment Comparison
n	Anzahl
n.k.	Nicht kalkuliert
NCL	Neuronale Ceroid-Lipofuszinose
NH	Natural History (Natürlicher Verlauf)
Non-R	Non-Responder
p	Probability (Wahrscheinlichkeit)
p190-201	Responder der Studie 190-201
p190-901	Responder der Studie 190-901
PP	Per Protokoll
PPT1	Palmitoyl-Protein-Thioesterase 1
PT	Preferred Term (Bevorzugter Begriff)
r	Reliability (Zuverlässigkeit)
R	Responder
RCT	Randomized Controlled Trial
RECORD	REporting of studies Conducted using Observational Routinely-collected Data
RR	Relatives Risiko
SAP	Statistischer Analyseplan
SAP ISE	Statistischer Analyseplan der „Integrated Summary of Efficacy“
SD	Standard Deviation (Standardabweichung)
SDO	Stable Dose Only (Stabile Dosierungsphase)
SE	Standarderror (Standardfehler)
SGB	Sozialgesetzbuch
SGB V	Sozialgesetzbuch Fünftes Buch
SMQ	Standardised MedDRA Query (Standardisierte MedDRA-Abfragen)

SOK	Systemorganklasse
STE	Surrogate Threshold Effects
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SUE	Schwerwiegendes Unerwünschtes Ereignis
TPP	Tripeptidylpeptidase
TPP1	Tripeptidylpeptidase 1
TREND	Transparent Reporting of Evaluations with Non-Randomized Design
TREND	Transparent Reporting of Evaluations with Nonrandomized Designs
u. a.	Unter Anderem
UE	Unerwünschtes Ereignis
UK	United Kingdom (Vereinigtes Königreich)
USA	United States of America
VerfO	Verfahrensordnung
vs.	<i>Versus</i>
WBV	Whole Brain Volume (Gesamtes Hirnvolumen)
WCMC	Weill Cornell Medical College
WHO	World Health Organization
z. B.	Zum Beispiel
ZNS	Zentralnervensystem

#### 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**

Die neuronale Ceroid-Lipofuszinose Typ 2 (CLN2) ist eine rasch voranschreitende, vererbte neurodegenerative Erkrankung mit gut vorhersagbarem Krankheitsverlauf von der Manifestation im späten Kindesalter bis zum Tod in der frühen Adoleszenz. In Deutschland liegt die Inzidenzrate bei 0,46 Erkrankungen pro 100.000 Lebendgeburten pro Jahr (siehe Modul 3).

Die Erstmanifestationen (Krämpfe und/oder Ataxie oft in Kombination mit einer verzögerten, frühen Sprachentwicklung) treten üblicherweise bei Kleinkindern im Alter von 2-4 Jahren (spätinfantil) auf [1,2]. Krampfanfälle wie generalisiert tonisch-klonische, partielle, myoklonische oder Absence-Epilepsien werden in der Regel behandlungsresistent [2–4]. Der weitere Verlauf der Erkrankung ist durch Ataxie, Bewegungsstörungen (Myoklonus, Dystonie und Chorea), und fortschreitende Demenz und Erblindung gekennzeichnet [2,3,5]. Die ersten schweren funktionalen Symptome treten im Alter von etwa drei Jahren auf. Sie sind gekennzeichnet durch den raschen Verlust sprachlicher und motorischer Fähigkeiten. In einem Zeitraum von etwa 2,5 Jahren gehen diese Fähigkeiten nach einem vorhersagbaren Muster vollständig verloren [5]. Die meisten Patienten werden bettlägerig, und benötigen zusätzliche Ernährungsmaßnahmen, da das Schluckvermögen durch die Erkrankung beeinträchtigt wird [3,6]. Eine künstliche Ernährung kann das Überleben verlängern [2]. Die durchschnittliche Lebenserwartung liegt bei 8-10 Jahren (siehe Abbildung 3) [1,5,7,8].

CLN2 wird durch Mutationen des CLN2-Gens verursacht, die zu einem Mangel des löslichen Enzyms Tripeptidylpeptidase 1 (TPP1) [5,9–11] führen. Die Abwesenheit der Enzymaktivität führt zu einer Akkumulation von abnormalem Speichermaterial innerhalb der Lysosomen von neuronalen Zellen, Gliazellen, und Retinazellen [9,12].

Bei Cerliponase alfa handelt es sich um eine rekombinate Form der humanen TPP1, welche zur Enzymersatztherapie („enzyme replacement therapy“, ERT) bei endogenem TPP1-Mangel angezeigt ist. Cerliponase alfa liegt als inaktives Proenzym (sogenanntes Zymogen) vor und wird innerhalb der Lysosomen aktiviert. Da Cerliponase alfa aufgrund der Molekülgröße die Blut-Hirn-Schranke nicht passiert [13], ist eine intrazerebrale Anwendung via intracerebroventrikulärem (ICV)-Port erforderlich. Der ICV-Port muss vor der ersten Infusion von Cerliponase alfa chirurgisch implantiert werden.

Cerliponase alfa wurde am 12.03.2015 gemäß der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 der Orphan Drug-Status zugesprochen [14]. Für Arzneimittel zur Behandlung eines seltenen Leidens („Orphan Drugs“) gilt gemäß § 35a SGB V Abs. 1 Satz 10 der medizinische Zusatznutzen durch die Zulassung als belegt [15]. Nach § 12 Nr. 1 der Verfahrensordnung (VerfO) ist lediglich das Ausmaß des Zusatznutzens für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, nachzuweisen [16]. Abgesehen von Cerliponase alfa, steht



derzeit weder in Europa noch weltweit eine allgemein akzeptierte oder zugelassene Therapie bei CLN2 zur Verfügung, die über unterstützende oder palliative Maßnahmen hinausgeht.

In Übereinstimmung mit § 12 Absatz 1 des 5. Kapitels der Verfo wird die folgende Fragestellung in dem vorliegenden Dossier untersucht: Wie groß ist das Ausmaß des Zusatznutzens von Cerliponase alfa unter Berücksichtigung patientenrelevanter Endpunkte zur Mortalität, Morbidität, Lebensqualität und Sicherheit in der zugelassenen Indikation? Zur Beantwortung der Fragestellung wird die Zulassungsstudie (einschließlich deren Extensionsstudie) verglichen mit Längsschnittdaten zum natürlichen Verlauf bei einem vergleichbaren Patientenkollektiv (“matched historical control”) herangezogen. Bei der historischen Kontrollgruppe handelt es sich weltweit um den bislang größten und vollständigsten Datensatz zu CLN2. Der Vergleich der Ergebnisse der Zulassungsstudie und der historischen Kontrolle ist auf diejenigen patientenrelevanten Endpunkte beschränkt, die in beiden Studien erhoben wurden.

### **Datenquellen**

Zur Bewertung des Ausmaßes des Zusatznutzens von Cerliponase alfa zur Behandlung eines seltenen Leidens sind die Zulassung und die Zulassungsstudien zu berücksichtigen. In Übereinstimmung mit § 12 der Verfo des G-BA ist der Zusatznutzen von Cerliponase zur Behandlung eines seltenen Leidens durch die Zulassung belegt [16].

Im vorliegenden Dossier werden die Ergebnisse der Zulassungsstudie 190-201/202 zur Bewertung herangezogen, und mit einer historischen Kontrolle unbehandelter Patienten mit vergleichbaren Charakteristika verglichen. Die Daten der historischen Kontrollgruppe (Studie 190-901) wurden im Rahmen des DEM-CHILD Patientenregisters erhoben, eines durch den FP7 Zuschuss der Europäischen Union (EU) geförderten Forschungsprogramms [17].

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

Bei Arzneimitteln für seltene Leiden sind die zulassungsrelevanten Studien bei der Nutzenbewertung zu berücksichtigen und es sind nur die Abschnitte 4.4.2 und 4.4.3 zu bearbeiten [18]. Somit entfällt die Verpflichtung zur Definition von Ein- und Ausschlusskriterien im Falle von Arzneimitteln zur Therapie seltener Leiden.

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

Aufgrund ihres „dramatischen“ Effekts wurde Cerliponase alfa bereits auf Basis der Phase 1/2-Studien 190-201/202 zugelassen.

Bei der relevanten Cerliponase alfa-Studie handelt es sich um eine einarmige Studie, da ein randomisiertes Design aus ethischen Gründen nicht möglich war. Eine Placebo-kontrollierte Studie war aufgrund der Notwendigkeit eines intrakraniellen, ventrikulären Katheters nicht vertretbar, da dieser für unbehandelte bzw. mit einem Placebo behandelte Patienten eine unakzeptable Belastung dargestellt hätte. Zudem hätte eine Placebo-Gruppe bedeutet, dass Patienten eine Behandlung vorenthalten wird, die diese dringend benötigen. Die Erkrankung wäre bei unbehandelten Patienten auch bei sehr kurzer Studiendauer deutlich und irreversibel

fortgeschritten. Eine aktiv-kontrollierte Studie war nicht möglich, da zur Behandlung der CLN2 keine vergleichbaren Therapien zur Verfügung stehen.

Entsprechend der Vorgaben evidenzbasierter Medizin wurde die nicht kontrollierte, prospektive Interventionsstudie 190-201/202 gemäß TREND-Statement berichtet (siehe Anhang 4-E) [19], dem derzeitigen Berichtsstandard für nicht randomisierte Interventionsstudien. Die Studie 190-901 zum natürlichen Verlauf wurde entsprechend des STROBE-Statements für epidemiologische Beobachtungsstudien dargestellt [20,21]. Das RECORD-Statement wurde hierbei berücksichtigt, welches sich auf Beobachtungsstudien auf Basis routinemäßig erhobener Gesundheitsdaten bezieht (siehe Anhang 4-E) [22].

Für Arzneimittel zur Behandlung seltener Leiden („Orphan drugs“) ist eine Bewertung des Verzerrungspotentials auf Studien- sowie Endpunktebene im Nutzendossier formal nicht erforderlich, da nur die Kapitel 4.4.2 und 4.4.3 des Dossiers zu bearbeiten sind. Aus Gründen der Transparenz wurde es dennoch als sinnvoll erachtet, andere Verzerrungsquellen der Studien und der Endpunkte zu bewerten, die einen Einfluss auf die Ergebnisse haben könnten. In Anhang 4-F wurden die Studie und patientenrelevanten Endpunkte hinsichtlich möglicher Verzerrungen untersucht. Auf eine Gesamtbewertung des Verzerrungspotentials wurde verzichtet, da dies nur bei randomisierten Studien erforderlich ist.

Den prospektiv definierten primären und sekundären Endpunkten wurden patientenrelevante Endpunkte, wie die Verbesserung des Gesundheitszustandes, die Verkürzung der Krankheitsdauer, die Verlängerung des Überlebens, die Verringerung von Nebenwirkungen oder die Verbesserung der Lebensqualität (QoL) zugeordnet [16].

Zur Bewertung des Ausmaßes des Zusatznutzens von Cerliponase alfa wird hinsichtlich des Fortschreitens der Erkrankung ein Vergleich der Studien 190-201/202 mit einer vergleichbaren („matched“) historischen Kontrollgruppe (Studie 190-901) herangezogen. Das Ausmaß des Zusatznutzens einer medizinischen Intervention kann auf Basis einer nicht kontrollierten Studie bewertet werden, vorausgesetzt, der Effekt der Intervention ist so groß („dramatisch“), dass dieser nicht durch Störgrößen erklärt werden kann. Es sollte gezeigt werden, dass die Intervention den (quasi-)deterministischen Krankheitsverlauf bei einer gut dokumentierten Serie von Patientinnen und Patienten umkehrt [23]. Zur Quantifizierung eines „dramatischen“ Effekts wurden die Qualität der Daten der historischen Kontrollgruppe analysiert und die Empfehlungen der GRADE-Arbeitsgruppe sowie des IQWiGs berücksichtigt (Nachweis von Effektgrößen von 5-10, ausgedrückt als relatives Risiko [RR] bei einem Signifikanzniveau von <1%) [20,23].

Da nur eine Studie mit einer fortlaufenden Extensionsphase zur Nutzenbewertung vorliegt, entfallen Meta-Analysen.

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

Basis der Bewertung des Zusatznutzens von Cerliponase alfa bildet die einarmige Studie 190-201/202, die der Zulassung von Cerliponase alfa zugrunde liegt. Die Studienergebnisse wurden mit einer historischen Kontrollgruppe aus Patienten mit ähnlichen Merkmalen („matched“)

verglichen (Analyse des natürlichen Krankheitsverlaufs: Studie 190-901). Tabelle 4-1 zeigt die patientenrelevanten Endpunkte, die zur Bewertung des Ausmaßes des Zusatznutzens von Cerliponase alfa herangezogen wurden.

Tabelle 4-1: Matrix der patientenrelevanten Endpunkte der eingeschlossenen Studie 190-201/202 mit dem zu bewertenden Arzneimittel

Patientenrelevante Zielgrößenkategorie	Endpunktparameter
Mortalität	Angesichts der Lebenserwartung der Patienten konnte der Endpunkt Mortalität in den Zulassungsstudien noch nicht erfasst werden. Die derzeit laufende Verlängerungsstudie 190-202 wird voraussichtlich erste Ergebnisse zur Mortalität liefern
Krankheitssymptome (Morbidität)	Anteil der Patienten mit patientenrelevanter Veränderung auf der krankheitsspezifischen 7-Punkte ML/HML Skala <sup>1</sup> (motorische Fähigkeiten und Sprachvermögen; primärer Endpunkt)
Gesundheitsbezogene Lebensqualität	Peds-QL (Fremdbeurteilung), CLN2-spezifische Skala (Fremdbeurteilung)
Nebenwirkungen	Gesamtzahl der unerwünschten Ereignisse (UEs) Gesamtzahl der UEs pro Systemorganklasse (SOK) Häufige UEs Schwerwiegende UEs (SUEs) und Todesfälle CTCAE („Common Terminology Criteria for Adverse Event“)  Grad $\geq 3$ UEs  Studien- oder Behandlungsabbruch aufgrund eines UEs UEs besonderen Interesses (AEOSI, „adverse event of special interest“): Überempfindlichkeits- und anaphylaktische Reaktionen (MedDRA SMQs, „standardised MedDRA query“) Infusionsassoziierte UEs (TRE, „temporally related event“) Erkrankungen des Nervensystems, MedDRA SMQ “Konvulsionen”, Status epilepticus, Hydrozephalus und Meningitis Unerwarteter und rascher, nicht anders erklärbarer Abfall des CLN2-Scores UEs, die mit dem ICV-Zugang in Zusammenhang standen.
<sup>1</sup> Zusammengesetzter Score aus zwei Skalen: M=motorische Fähigkeiten, L=Sprachvermögen (Language). Jede Einzelskala reicht von 0 (schlechtester Wert) bis 3 (bester Wert). Die kombinierte 7-Punkte Skala umfasst die Kategorien 0 bis 6.	

### ***Krankheitsprogression anhand der krankheitsspezifischen ML/HML-Skala***

Um die Krankheitsprogression in der pivotalen klinischen Studie und in der historischen Kontrollgruppe zu erfassen wurden die zwei Domänen für Motor- und Sprachfunktion der etablierten CLN2 Rating-Skala eingesetzt. Die beiden Domänen der Rating-Skala erfassen den Verlauf der Erkrankung in der schnellen Progressionsphase am verlässlichsten [5]. Für jede Einzeldomäne wird der Funktionsstatus auf einer Skala von 3 (normale Funktion) bis 0

(kompletter Verlust der Funktion) erhoben. Die kombinierte Skala für Motor- und Sprachfunktion ist eine 7-Punkte-Skala mit den Werten von 0-6 [1,24].

In der historischen Vergleichsstudie 190-901 wurde für die Erhebung der Motor- und Sprachfunktionen die ursprünglich entwickelte Skala verwendet [5], die auch für die Dokumentation des natürlichen Verlaufs der Erkrankung im Universitätsklinikum Hamburg Eppendorf eingesetzt wurde (auf diese kombinierte Skala zur Erhebung der Motor- und Sprachfunktionen wird im Folgenden als „*HML-Skala*“ Bezug genommen). Für die Studie 190-201 wurde diese Skala in Kollaboration mit den Entwicklern der HML-Skala minimal angepasst, um die Kategorien der Skalen auf Basis der historischen Kriterien für eine multizentrische Studie zu präzisieren (auf diese Skala wird im Folgenden als „*ML-Skala*“ Bezug genommen). Ein Vergleich der beiden Skalen ist in Tabelle 4-29 aufgeführt.

Die mittels ML/HML-Skala ermittelte relevante Änderung stellte den primären Endpunkt für den Vergleich mit der historischen Kontrollgruppe dar. Eine Änderung im primären Endpunkt um zwei Punkte war prospektiv als Schwellenwert definiert, um Patienten zu definieren, die auf die Therapie ansprechen (<2 Punkte Verschlechterung) oder nicht ( $\geq 2$  Punkte Verschlechterung). Dieses Kriterium entspricht der durchschnittlichen Veränderung in einem Jahr während der Phase der schnellen Krankheitsprogression (siehe Kapitel 4.4.2.3.1.1). Zudem entspricht eine 2-Punkte-Veränderung einem klinisch relevanten Unterschied, der für die betroffenen Patienten von hoher Relevanz ist. Eine 2-Punkte-Veränderung liegt zudem deutlich über einer mittels statistischer Verteilungsmethoden abgeleiteten Minimal Important Difference (MID) (siehe Kapitel 4.4.2.3.1.2).

Tabelle 4-2 gibt einen Überblick der Ergebnisse der Ereigniszeit- und Responder-Analysen für den primären Endpunkt, jeweils unter Einsatz des vordefinierten 2-Punkte-Schwellenwerts. Diese vergleichenden Analysen sind für die Nutzenbewertung relevant, da sie einen für die Patienten klinisch bedeutsamen Unterschied erfassen. Die Ergebnisse belegen - verglichen mit der historischen Kontrollgruppe (1:1 Paarbildung, „matched“ control) - einen „dramatischen“ Effekt von Cerliponase alfa. Zusätzlich sind in Tabelle 4-2 Responder-Analysen (1:1 Paarbildung) getrennt für die Motordomäne sowie Sprachdomäne dargestellt, wobei bei diesen Analysen eine Änderung um 1 Punkt als Erfolgskriterium galt.

Verglichen mit der unbehandelten historischen Kontrollgruppe stabilisierte die Behandlung mit Cerliponase alfa die Erkrankung unmittelbar nach Therapiebeginn. In der Kaplan-Meier-Analyse war das Risiko eines Fortschreitens der Erkrankung (gemessen als Abfall des ML/HML-Scores um  $\geq 2$  Punkte oder Score von Null) durch die Behandlung um das 10-fache reduziert (HR=0,10 [95% KI: 0,03; 0,38], p=0,0005). Bei Patienten der historischen Kontrolle nahm der Wert auf der HML-Skala innerhalb von 285 Tagen im Median um zwei Punkte ab. Demgegenüber konnte der Median bei den behandelten Patienten nicht ermittelt werden, da die Erkrankung nicht fortgeschritten war. Dieser Behandlungseffekt von Cerliponase alfa ist, verglichen mit der unbehandelten historischen Kontrolle, als „dramatisch“ einzustufen (Tabelle 4-2).

Der Anteil der Patienten in der Cerliponase alfa-Gruppe, die auf die Behandlung angesprochen hatten (Responder, definiert als Abfall des ML/HML-Scores um  $<2$  Punkte innerhalb von 48 Wochen), war signifikant höher als in der historischen Kontrollgruppe („matched“; RR=2,00 [95% KI: 1,2420; 3,2207],  $p=0,0044$ ). Hinsichtlich des Anteils der Non-Responder (definiert als Abfall des ML/HML-Scores um  $\geq 2$  Punkte innerhalb von 48 Wochen) betrug das RR 0,16667 ([95% KI 0,04212; 0,65942],  $p=0,01067$ ; Kehrwert RR: 6,00) zugunsten von Cerliponase alfa verglichen mit der historischen Kontrollgruppe („matched“). Diese Ergebnisse unterstützen den „dramatischen“ Behandlungseffekt von Cerliponase alfa aus den Ereigniszeitanalysen (Tabelle 4-2).

Auch die Responder-Analysen der einzelnen Domänen der ML/HML-Skala belegen einen hochsignifikanten Behandlungseffekt; bei diesen Analysen galt eine Änderung um einen Punkt als Erfolgskriterium. Die RR für Non-Response auf der Sprachskala betrug 0,1250 ([95% KI 0,327; 0,4773],  $p=0,0023$ ; Kehrwert: 8). Die motorische Komponente unterschied sich zwischen den Gruppen etwas weniger deutlich, dennoch war der Unterschied hochsignifikant zugunsten von Cerliponase alfa (RR 1,900 [95% KI 1,1676; 3,0919],  $p=0,0098$ ) (Tabelle 4-2). Diese Ergebnisse zeigen, dass Cerliponase alfa positive Effekte auf unterschiedliche physiologische Funktionen sowie unterschiedliche anatomische Regionen innerhalb des Zentralnervensystems ausübt, die durch die ML/HML-Skala erfasst werden können (Tabelle 4-2).

Tabelle 4-2: Zusammenfassung der Ergebnisse für Krankheitsprogression gemessen mittels ML/HML-Skala: Cerliponase alfa vs. historische Kontrolle nach 1:1 Paarbildung entsprechend klinisch relevanter prognostischer Faktoren

Endpunkt	190-201/190-901 (1:1 Paarbildung)			Bewertung des Effekts
	Effektmaß	Effekt	p-Wert	
<i>ML/HML Skala / Primäre Analysen (siehe Kapitel 4.4.2.3.1.3)</i>				
Zeit bis zur ersten stabilen Abnahme des ML/HML-Scores um $\geq 2$ Punkte oder ML = 0	HR, Kaplan-Meier	0,10 95% KI 0,03; 0,38	0,0005	Schwerwiegendes Krankheitssymptom; 10-fach höheres Risiko einer Krankheitsprogression bei unbehandelten Patienten; erheblicher Zusatznutzen
Responder (Abnahme des ML/HML-Scores um $<2$ Punkte/48 Wochen)	RR	2,00 95% KI 1,2420; 3,2207	0,00435	Schwerwiegendes Krankheitssymptom; RR (iRR) $>5$ („dramatischer“ Effekt); erheblicher Zusatznutzen
Non-Responder (stabile Abnahme des ML/HML-Scores um $\geq 2$ Punkte/48 Wochen)	RR	0,16667 95% KI 0,04212; 0,65942 (iRR: 6,00)	0,01067	

Endpunkt	190-201/190-901 (1:1 Paarbildung)			Bewertung des Effekts
	Effektmaß	Effekt	p-Wert	
<i>Einzeldomänen (Motorische Fähigkeiten, Sprachvermögen) (siehe Kapitel 4.4.2.3.1.5)</i>				
Responder (Abnahme in der motorischen Domäne um <1 Punkt/48 Wochen)	RR	1,900 95% KI 1,1676; 3,0919	0,0098	
Non-Responder (stabile Abnahme in der motorischen Domäne um $\geq 1$ Punkt/48 Wochen)	RR	0,250 95% KI 0,0817; 0,7652 (iRR: 4,00)	0,0151	
Responder (Abnahme in der Sprachdomäne um <1 Punkt/48 Wochen)	RR	3,800 95% KI 1,7461; 8,2696	0,0008	
Non-Responder (stabile Abnahme in der Sprachdomäne um $\geq 1$ Punkt/48 Wochen)	RR	0,1250 95% KI 0,327; 0,4773 (iRR: 8,00)	0,0023	
iRR: inverses relatives Risiko; KI: Konfidenzintervall; RR: relatives Risiko Kriterien zur Paarbildung (matching): Identischer ML/HML-Skalenwert und Altersunterschied $\leq 12$ Monate zu Baseline (300 mg-Baseline in Studie 190-201).				

Cerliponase alfa zeigt auf der krankheitsspezifischen ML/HML-Skala über 48 Wochen einen „dramatischen“ Behandlungseffekt, der langfristig erhalten bleibt. Dies wird auch durch die Ergebnisse von Sensitivitätsanalysen gestützt, die auf Alter, Krankheitsprogression und Genotyp kontrollieren. Darüberhinaus zeigen die Follow-up-Daten nach 96 Wochen, dass der Unterschied zwischen den behandelten und unbehandelten Patienten hinsichtlich des Endpunkts Krankheitsprogression über 96 Wochen noch zunimmt (HR=0,05 [95%KI: 0,01; 0,17],  $p < 0,0001$ ) (siehe Kapitel 4.4.2.3.1.6). Hervorzuheben ist, dass Cerliponase alfa die Erkrankung innerhalb kurzer Zeit stabilisiert bzw. die Krankheitsprogression „dramatisch“ verlangsamt, die ansonsten rasch und nach einem vorhersagbaren Muster voranschreitet (gemessen unter Einsatz einer CLN2-spezifischen, klinischen Bewertungsskala).

### **MRI Analysen**

Quantitative Volumenmessungen zeigten eine Stabilisierung des Verlustes der grauen Substanz bis Woche 49 nach Behandlungsbeginn mit Cerliponase alfa. Im Vergleich zu einer prospektiven historischen Kohorte (unbehandelte Patienten), in welcher der Populationsmittelwert des Volumenverlustes bei 14,5% gemessen wurde [25], lag der Mittelwert bei behandelten Patienten bei 9,7% und blieb danach nahezu unverändert [26,27]. Interpretationen einer Korrelation von klinischen Messungen mit den anatomischen Veränderungen sind zum jetzigen Zeitpunkt jedoch mit Vorsicht zu führen, da das Volumen der grauen Substanz von behandelten Patienten nicht nur aufgrund der Krankheitsprogression abnehmen könnte, sondern auch durch andere Faktoren wie durch den therapeutischen Effekt

(Abbau des abnormen lysosomalen Speichermaterials). Für eine Validierung dieses Endpunktparameters sind folglich noch weitere Forschungsarbeiten erforderlich.

### *QoL*

Es wurde erwartet, dass die gesundheitsbezogene Lebensqualität bei unbehandelten Patienten parallel zur raschen Krankheitsprogression deutlich abnimmt. Jedoch liegen für den Endpunkt gesundheitsbezogene Lebensqualität (gemessen mittels PedsQL™) keine historischen Kontrolldaten vor. Daher werden für diesen Endpunkt deskriptive Längsschnittdaten aus der klinischen Studie für die behandelten Patienten dargestellt.

Die Ergebnisse zur Lebensqualität sind in Tabelle 4-3 und Tabelle 4-4 dargestellt. Die Lebensqualität, ausgedrückt als Summenscore der beiden PedsQL™-Module (sowohl Elternbericht für Kleinkinder als auch familiäre Belastung), besserte sich im Studienverlauf numerisch.

Der Mittelwert ( $\pm$ SD) des Summenscores des PedsQL™/Elternbericht für Kleinkinder nahm zwischen Baseline und Woche 49 der stabilen Dosisphase um 2,6 (12,16) Punkte zu, was einer Verbesserung um etwa 4.3% entspricht (ITT, gesamtes Dosisintervall). Über denselben Zeitraum nahm die Lebensqualität gemessen mittels PedsQL™/Familiäre Belastungen im Mittel (SD) um 3,7 (19,04) Punkte zu, d. h. der Score verbesserte sich um etwa 6.0% (Tabelle 4-3).

Auch der CLN2-spezifische QoL-Score nahm zwischen Baseline und Woche 49 der stabilen Dosisphase im Mittel (SD) um 8,1 (14,33) Punkte zu, entsprechend einer Verbesserung von 10,9%. Der mittlere Wert ( $\pm$ SD) der Subdomäne Krampfaktivität („seizures“) verbesserte sich um 18,2 (23,02) Punkte. Die Behandlung mit Cerliponase alfa hatte keinen negativen Effekt auf irgendeine der Einzeldomänen der CLN2-QoL-Skala (der geringfügige Abfall um 0,8 Punkte hinsichtlich üblicher Alltagsaktivitäten wird als klinisch nicht relevant eingestuft) (Tabelle 4-4).

Tabelle 4-3: Überblick der Ergebnisse zur Lebensqualität (PedsQL™; ITT, n=23)

<b>Endpunkt</b>	<b>Änderung bis zur letzten Erhebung (Punkte<sup>1</sup>), Mittelwert (SD)</b>
PedsQL™ Fremdbericht bei Kleinkindern (generischer Gesamtscore; Tabelle 4-54)	2,6 (12,16)
Körperliche Fähigkeiten	-6,1 (21,72)
Emotionale Fähigkeiten	7,3 (26,22) (n=22)
Soziale Fähigkeiten	7,8 (16,22)
Schule	6,3 (29,82)
Psychosozialer Summenscore	7,9 (14,91)
PedsQL™ „Family Impact Module“, Fremdbericht, Total Score (Tabelle 4-54)	3,7 (19,04)
Körperliche Funktion	1,6 (20,86)
Emotionale Fähigkeiten	10,2 (22,99)
Soziale Fähigkeiten	-4,6 (25,23)
Kognitive Fähigkeiten	7,4 (32,82)
Kommunikative Fähigkeiten	2,2 (28,00)
Summenscore (körperlich, emotional, sozial und kognitiv)	4,0 (19,10)
<sup>1</sup> Ein Score von 0 entspricht dem schlechtesten Wert und ein Score von 100 dem besten Wert.	

Tabelle 4-4: Überblick der Ergebnisse zur Lebensqualität (CLN2-spezifisches QoL-Instrument; ITT, n=23)

<b>Endpunkt</b>	<b>Änderung bis zur letzten Erhebung (Punkte), Mittelwert (SD)</b>
CLN2-spezifische Lebensqualität, Gesamtscore (Tabelle 4-56)	8,1 (14,33)
Krampfanfälle	18,2 (23,02)
Ernährung (ohne Gastrostomiesonde)	9,8 (22,15) (n=21)
Schlaf	4,1 (19,74)
Verhalten	7,2 (18,01)
Alltagsaktivitäten	-0,8 (18,10)
Ein Score von 0 entspricht dem schlechtesten Wert und ein Score von 100 dem besten Wert.	

In Anbetracht des rasch voranschreitenden Krankheitsverlaufs und der außerordentlich schlechten Prognose unbehandelter Patienten repräsentiert eine Stabilisierung/Verbesserung der Gesamtscores, die sich bei zwei QoL-Instrumenten unabhängig voneinander zeigt, einen bedeutsamen Nutzen für die Patienten (sowie deren Eltern/Familien).



### ***Rate unerwünschter Ereignisse***

Für den Endpunkt „Unerwünschte Ereignisse“ (UE) liegen keine historischen Kontrolldaten vor. Daher werden die Inzidenzen der in der Studie 190-201/202 aufgetretenen UEs dargestellt. Die Ergebnisse werden qualitativ bewertet und der Behandlungseffekt von Cerliponase alfa sollte bei der Interpretation berücksichtigt werden.

*Gesamtzahl der UEs; häufigste, berichtete Systemorganklassen (SOK); häufigste UEs („preferred terms“, PTs); schwerwiegende UEs (SUE) und UEs der Grade  $\geq 3$*

Die hier berichteten Ergebnisse bilden den aktuellsten Datenschnitt vom 3. Juni 2016 der Folgestudie 190-202 ab (gesamtes Dosisintervall einschließlich der Studie 190-201; bei allen in der Studie 190-201/202 mit einer Dosis von 300 mg behandelten Patienten betrug die Exposition mindestens 72 Wochen bei einer Dosis von 300 mg, ausgenommen ein Patient, der die Studie nach nur einer Dosis abgebrochen hatte).

Alle Patienten (n=24/24, 100%) entwickelten mindestens ein UE.

Die meisten UEs traten innerhalb der SOKs „Erkrankungen des Nervensystems“ (100%), „Infektionen und parasitäre Erkrankungen“ (100%), „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ (79%), „Erkrankungen des Gastrointestinaltrakts“ (79%), „Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen“ (79%) und „Psychiatrische Erkrankungen“ (50%) auf.

Fieber (71%), Krampfanfälle (58%), Erbrechen (50%), Infektionen der oberen Atemwege (50%), Epilepsie (46%), generalisierte, tonisch-klonische Anfälle (46%) und Rhinitis (42%) stellten sehr häufige UEs dar.

Bis zum Datenschnitt im Juni 2016 traten bei 19 der 24 Patienten (79%) insgesamt 51 schwerwiegende UEs (SUEs) auf, viele davon nur einmalig. Todesfälle wurden während der Studien 190-201/202 nicht berichtet. Insgesamt 12 der 24 Patienten (50%) entwickelten ein schweres CTCAE Grad 3-UE und ein Patient (4%) ein Grad 4-UE.

### *UEs besonderen Interesses*

Eine standardisierte MedDRA-Abfrage (SMQ) nach anaphylaktischen Reaktionen erbrachte kein Ergebnis. Bis Juni 2016 traten bei 14 Patienten (58%) insgesamt 36 Überempfindlichkeitsreaktionen auf. Die folgenden UEs (PT) wurden durch eine breite MedDRA-Abfrage (SMQ „Hypersensibilität“) identifiziert: Überempfindlichkeit (n=9, 38%, 15 UEs), jahreszeitbedingte Allergie (n=1, 4%, 2 UEs), Konjunktivitis (n=4, 17%, 5 UEs), Ausschlag (n=2, 8%, 7 UEs), Urtikaria (n=2, 8%, 2 UEs), Dermatitis (n=1, 4%, 1 UE), Kontaktdermatitis (n=2, 8%, 3 UEs) und Stomatitis (n=1, 4%, 1 UE). Nur drei Patienten (13%) entwickelten eine CTCAE Grad 3-Überempfindlichkeit.

Bis zum Datenschnitt Juni 2016 entwickelten alle Patienten (n=24, 100%) mindestens ein UE, welches in zeitlichem Zusammenhang mit der Behandlung stand („temporally-related event“, TRE). Die häufigsten dieser UEs (PT) waren Fieber (46%), Überempfindlichkeit (38%),

Krampfanfall (33%), Erbrechen (29%) und Tremor (25%). CTCAE Grad 3-TRE traten bei 7 Patienten (29%) auf.

Alle Patienten (n=24/24, 100%) entwickelten mindestens ein UE innerhalb der SOK „Erkrankungen des Nervensystems“, wovon 83% geringer oder mäßig schwerer Ausprägung waren (CTCAE Grad 1 oder 2). Insgesamt 362 UEs bei 23 Patienten (96%) wurden durch eine standardisierte MedDRA-Abfrage nach Konvulsionen identifiziert. Als häufigste Ereignisse wurden dabei Epilepsie (175 Ereignisse bei 11 Patienten, 46%) und Krampfanfälle (130 Ereignisse bei 7 Patienten, 29%) berichtet.

Bis zum Datenschnitt Juni 2011 traten bei 11 der 24 Patienten (46%) 30 UEs auf, die mit dem ICV-Zugang oder den hierzu eingesetzten Medizinprodukten in Zusammenhang standen.

Das Ausmaß des Zusatznutzens (oder Schadens) von Cerliponase alfa ist hinsichtlich der Sicherheit und Verträglichkeit nicht quantifizierbar, da bei Patienten mit CLN2 keine Daten zu UEs unter üblichen Versorgungsbedingungen vorliegen und ein historischer Vergleich daher nicht möglich ist. Die Sicherheit und Verträglichkeit von Cerliponase alfa wurde qualitativ in Relation zum Nutzen der Behandlung bewertet. Dabei ergeben sich keine Hinweise auf einen nicht vertretbaren Schaden durch die Behandlung mit Cerliponase alfa, wenn dieser dem „dramatischen“ Effekt auf die Krankheitsprogression gegenübergestellt wird. Bei unbehandelten Patienten ist eine Progression irreversibel und mit sehr schwerwiegenden Symptomen verbunden. Bei nahezu allen UEs in Zusammenhang mit Cerliponase alfa handelte es sich um UEs, die bei einer ERT zu erwarten sind [28].

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

Alle Auswertungen der ML/HML-Skala, die sich auf einen Vergleich zwischen den mit Cerliponase alfa (300 mg alle zwei Wochen) behandelten Patienten der Studie 190-201/202 und der unbehandelten, historischen Kontrolle der Studie 190-901 beziehen, zeigen einen deutlichen und statistisch hoch signifikanten Effekt zugunsten von Cerliponase alfa. Die Ergebnisse einschließlich deskriptiver Analysen der Änderung im Studienverlauf, Responder-Analysen, Ereigniszeitanalysen (Zeit bis zur dauerhaften Krankheitsprogression um  $\geq 2$  Punkte auf der ML/HML-Skala) sowie der Rate, mit der die Erkrankung fortschreitet, belegen übereinstimmend, dass die Erkrankung bei behandelten Patienten im Gegensatz zu unbehandelten Patienten stabil bleibt. Bei der primären Wirksamkeitsanalyse (Ereigniszeit- sowie Responder-Analysen) wurden patientenrelevante Schwellenwerte zur Bewertung des Behandlungserfolgs herangezogen; diese belegen einen „dramatischen“ Behandlungseffekt von Cerliponase alfa. Derartig große Behandlungseffekte sind durch Störfaktoren nicht erklärbar und erlauben daher eine Quantifizierung des Zusatznutzens auf Basis eines historischen Vergleichs.

Alle Ergebnisse, die für die Nutzenbewertung relevant sind, werden durch verschiedene Sensitivitätsanalysen (unterschiedliche Auswertungspopulationen und -verfahren, Kriterien zur Paarbildung zwischen unbehandelten, historischen Kontrollen und behandelten Patienten sowie Ausgangspunkts) bestätigt und diese belegen die Robustheit der Ergebnisse. Die Ergebnisse

unter Einsatz der vorab definierten Paarbildungskriterien sind denjenigen ohne Paarbildung vergleichbar und jede der Analysen stützt die primäre Wirksamkeitsanalyse. Die Kriterien zur Paarbildung zwischen den Teilnehmern der Studie 190-201 und den Patienten der historischen Kontrollgruppe entsprechen den Haupteinschlusskriterien der Studie 190-201 und stellen prognostisch wichtige Faktoren dar.

Der Verlust bereits erworbener Fähigkeiten ist bei unbehandelten Patienten mit CLN2 irreversibel und die Erkrankung verschlechtert sich bei diesen Patienten innerhalb eines Jahres um etwa zwei Punkte (gemessen auf der ML/HML-Skala). Die Belastung durch die Erkrankung ist sowohl für die Patienten als auch Familien, die sich um Kinder mit CLN2 kümmern, enorm, und auch die psychische Belastung ist keinesfalls zu unterschätzen, die mit einer tödlich verlaufenden Erkrankung im Kindesalter einhergeht. Derzeit stehen nur symptomatische oder palliative Therapien zur Behandlung der CLN2 zur Verfügung. Demgegenüber zeigt die Behandlung mit Cerliponase alfa einen langanhaltenden Therapieeffekt: Während die Erkrankung bei den Patienten der historischen Kontrollgruppe stetig und nahezu gleichförmig voranschreitet, bleibt die Erkrankung unter Behandlung mit Cerliponase alfa über die gesamte Dauer der Nachverfolgung (mindestens 96 Wochen) stabil oder bessert sich bei einigen Patienten sogar. Hervorzuheben ist, dass die *a priori* definierten Sekundäranalysen der 9-Punkt- (einschließlich Sehvermögen) und 12-Punkt-Skala (zusätzlich Krampfanfälle) eine weitreichende Stabilisierung der Erkrankung auf allen untersuchten Ebenen zeigen, d. h. die positiven Effekte von Cerliponase alfa sind nicht nur auf motorische und sprachliche Fähigkeiten beschränkt. Die Ergebnisse der MRI-Analysen unterstützen die klinischen Ergebnisse zur Stabilisierung der Erkrankung auf der anatomischen Ebene.

Zur Erfassung der Lebensqualität wurde der PedsQL™-Fragebogen (Eltern-Fremdbericht und familiäre Belastung) sowie eine CLN2-spezifische Skala eingesetzt, die Biomarin basierend auf den Angaben betroffener Eltern entwickelt hat. Die Lebensqualität blieb unter Cerliponase alfa bemerkenswert stabil und es ergaben sich keinerlei Anzeichen einer Verschlechterung. Im Gegenteil zeigten sich bei einigen Domänen positive Trends im Sinne einer höheren Lebensqualität. Der Zusatznutzen von Cerliponase alfa ist allerdings hinsichtlich der Lebensqualität formal nicht quantifizierbar, da entsprechende Daten bei den Patienten der historischen Kontrollgruppe nicht vorliegen.

In Anbetracht der außerordentlichen Schwere der Erkrankung und des „dramatischen“ Behandlungseffekts von Cerliponase alfa sind lediglich schwere, schwerwiegende oder lebensbedrohliche UEs sowie UEs, die zu dauerhaften Beeinträchtigungen führen, geeignet, das Nutzen-Risiko-Verhältnis zu ändern. Die in der klinischen Studie aufgetretenen Überempfindlichkeitsreaktionen und begleitenden Symptome werden als besonders relevant eingestuft. Die schweren und schwerwiegenden UEs waren beherrschbar und klangen ohne Folgeerscheinungen ab. Keiner der eingeschlossenen Patienten brach die Therapie aufgrund eines UEs ab.

In striktem Einklang mit den Kriterien des § 5 Absatz 7 des 5. Kapitels der Verfo (3) wird das Ausmaß des Zusatznutzens von Cerliponase alfa zur Behandlung von Patienten mit CLN2 als **erheblich** bewertet. Mit Cerliponase alfa steht erstmals eine Therapie zur Verfügung, die eine

langfristige Freiheit von schwerwiegenden Symptomen bewirkt, speziell der für die Patienten bedeutsamen, irreversiblen Krankheitsprogression, die unbehandelt kontinuierlich fortschreitet und mit dauerhaften, schwerwiegenden Symptomen einhergeht. Die Nebenwirkungen von Cerliponase alfa sind gut beherrschbar und hinsichtlich Inzidenz und Schwere akzeptabel, da die Lebensqualität der Patienten unter Therapie erhalten bleibt.

## 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).

*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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

### 4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung

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

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

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

### **4.2.3 Informationsbeschaffung**

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

#### **4.2.3.1 Studien des pharmazeutischen Unternehmers**

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

#### **4.2.3.2 Bibliografische Literaturrecherche**

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

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

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

Die bibliografische Literaturrecherche soll mindestens in den Datenbanken MEDLINE und EMBASE sowie in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials

(Clinical Trials)“ durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Datenbanken (z. B. CINAHL, PsycINFO etc.) durchgeführt werden.

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

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

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **4.2.3.3 Suche in Studienregistern**

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

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

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

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

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

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

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **4.2.3.4 Selektion relevanter Studien**

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

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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>1</sup>. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer

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



Beobachtungsstudien soll mindestens den Anforderungen des TREND-<sup>2</sup> bzw. STROBE-Statements<sup>3</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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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

<sup>2</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>3</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.

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<sup>4</sup> erfolgen. In begründeten Ausnahmefällen sollen zusätzlich Modelle mit festen Effekten eingesetzt werden. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, sollen diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt werden.

Für kontinuierliche Variablen soll die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges'  $g$ , als Effektmaß eingesetzt werden. Bei binären Variablen sollen Meta-Analysen primär sowohl anhand des Odds Ratios als auch des Relativen Risikos durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen soll ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet<sup>5</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 des Maßes  $I^2$  und des statistischen Tests auf Vorliegen von Heterogenität<sup>6</sup> erfolgen. Ist die Heterogenität der Studienergebnisse nicht bedeutsam, soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten Ausnahmefällen gepoolt werden. Außerdem soll untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.2.5.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.2.5.5).

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

Nicht zutreffend

#### **4.2.5.4 Sensitivitätsanalysen**

Zur Einschätzung der Robustheit der Ergebnisse sollen Sensitivitätsanalysen hinsichtlich methodischer Faktoren durchgeführt werden. Die methodischen Faktoren bilden sich aus den im Rahmen der Informationsbeschaffung und -bewertung getroffenen Entscheidungen, zum Beispiel die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des

---

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

<sup>5</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>6</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.*

Nicht zutreffend.

#### **4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren**

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

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

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

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

*Benennen Sie die durchgeführten Subgruppenanalysen. Begründen Sie die durchgeführten Subgruppenanalysen bzw. die Untersuchung von Effektmodifikatoren oder den Verzicht auf*

*solche Analysen. Beschreiben Sie die für diese Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.*

Nicht zutreffend.

#### 4.2.5.6 Indirekte Vergleiche

Zurzeit sind international Methoden in der Entwicklung, um indirekte Vergleiche zu ermöglichen. Nicht adjustierte indirekte Vergleiche (d. h. Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen Komparator) stellen dabei keine valide Analyseverfahren dar, der Einsatz einfacher adjustierter indirekter Vergleiche ist möglich<sup>7</sup>. Komplexe Verfahren für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche werden in der Literatur unterschiedlich bezeichnet, z. B. als „Mixed-Treatment-Comparison(MTC)-Meta-Analysen“<sup>8</sup>, „Multiple-Treatment-Meta-Analysen“<sup>9</sup> oder auch „Netzwerk-Meta-Analysen“<sup>10</sup>, sie gehen aber im Prinzip von denselben wesentlichen Annahmen aus.

Grundannahme für solche komplexen Analysen ist die Annahme der Konsistenz innerhalb des zu analysierenden Netzwerkes. Als Inkonsistenz wird dabei die Diskrepanz zwischen dem Ergebnis eines direkten und eines oder mehreren indirekten Vergleichen verstanden, die nicht mehr nur durch Zufallsfehler oder Heterogenität erklärbar ist<sup>11</sup>.

Da das Ergebnis eines indirekten Vergleichs maßgeblich von der Auswahl des Brückenkomparators bzw. der Brückenkomparatoren abhängen kann, ist die Wahl des Brückenkomparators bzw. der Brückenkomparatoren zu begründen. Dies gilt insbesondere dann, wenn eine Beschränkung auf ein oder mehrere Brückenkomparatoren vorgenommen wird, obwohl Daten zu anderen Therapieoptionen, die ebenfalls als Brückenkomparatoren in Frage kommen, vorliegen.

Insgesamt ist es notwendig, die zugrunde liegende Methodik genau und reproduzierbar zu beschreiben und die Annahme der Konsistenz zu untersuchen<sup>12</sup>.

---

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

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

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

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

<sup>11</sup> B. Schöttker, D. Lüthmann, D. Boulkhemair, and H. Raspe. Indirekte Vergleiche von Therapieverfahren. *Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.*

<sup>12</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.

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

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

Nicht zutreffend.

### 4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen

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

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

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

#### 4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel

##### 4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel

###### 4.3.1.1.1 Studien des pharmazeutischen Unternehmers

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

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

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

Tabelle 4-5: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer	Therapiearme
<Studie 1>	ja	ja	abgeschlossen	12 Monate	Medikament A, Medikament B, Placebo

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 Dossiers maßgeblichen Zeitpunkt liegen.

Nicht zutreffend.

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

#### 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.*



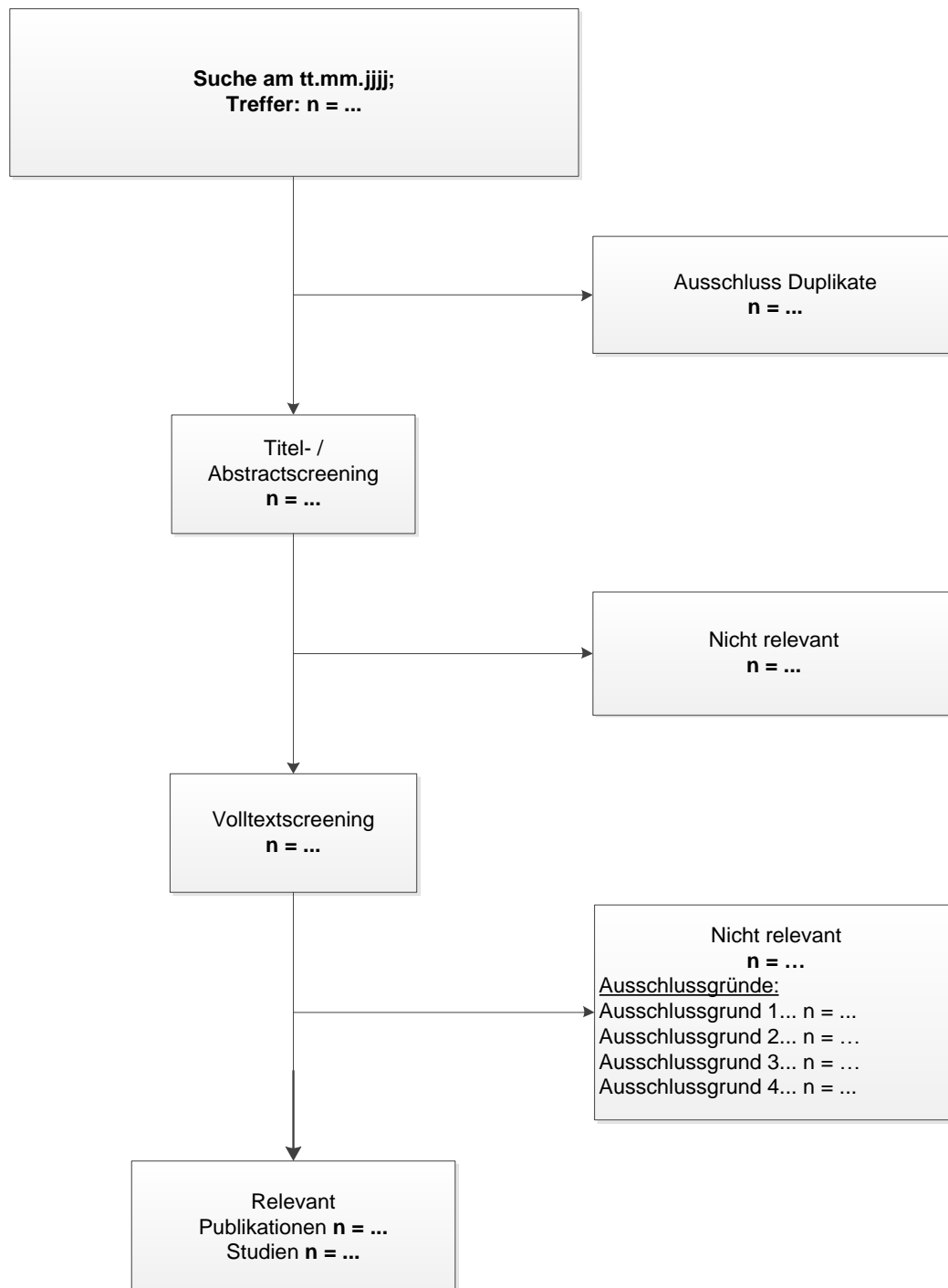


Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel

#### 4.3.1.1.3 Studien aus der Suche in Studienregistern

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern identifiziert wurden. Geben Sie dabei an, in welchem Studienregister die Studie identifiziert wurde und welche Dokumente dort zur Studie jeweils hinterlegt sind (z. B.

*Studienregistereintrag, Bericht über Studienergebnisse etc.). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-5) 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 – RCT mit dem zu bewertenden Arzneimittel

Studie	Identifikationsorte (Name des Studienregisters 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)
<Studie 1>	clinicaltrials.gov [6, 7]	ja	nein	abgeschlossen

a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.

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

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel**

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

*Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Quellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Hierbei sollen die Studien durch Zwischenzeilenüberschriften ggf. sinnvoll*

angeordnet werden, beispielsweise nach Therapieschema (Akut-/Langzeitstudien) und jeweils separat nach Art der Kontrolle (Placebo, zweckmäßige Vergleichstherapie, beides). Sollten Sie eine Strukturierung des Studienpools vornehmen, berücksichtigen Sie diese auch in den weiteren Tabellen in Modul 4.

Tabelle 4-8: 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)	Studienbericht (ja/nein [Zitat])	Registereintrag <sup>c</sup> (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
<b>ggf. Zwischenüberschrift zur Strukturierung des Studienpools</b>						
<b>placebokontrolliert</b>						
<Studie 1>	ja	ja	nein	ja [5]	ja [6, 7]	ja [8]
<b>aktivkontrolliert, zweckmäßige Vergleichstherapie(n)</b>						
<p>a: Bei Angabe „ja“ sind jeweils die Zitate der Quelle(n) (z. B. Publikationen, Studienberichte, Studienregister-einträge) mit anzugeben, und zwar als Verweis auf die in Abschnitt 4.7 genannte Referenzliste. Darüber hinaus ist darauf zu achten, dass alle Quellen, auf die in dieser Tabelle verwiesen wird, auch in Abschnitt 4.6 (Liste der eingeschlossenen Studien) aufgeführt werden.</p> <p>b: Studie, für die der Unternehmer Sponsor war.</p> <p>c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p>						

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

##### 4.3.1.2.1 Studiendesign und Studienpopulationen

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

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

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

Tabelle 4-9: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

<b>Studie</b>	<b>Studiendesign</b> <RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	<b>Population</b> <relevante Charakteristika, z. B. Schweregrad>	<b>Interventionen</b> (Zahl der <b>randomisierten</b> <b>Patienten</b> )	<b>Studiendauer</b> <ggf. Run-in, Behandlung, Nachbeobachtung>	<b>Ort und Zeitraum</b> <b>der Durchführung</b>	<b>Primärer</b> <b>Endpunkt;</b> <b>patientenrelevante</b> <b>sekundäre</b> <b>Endpunkte</b>
<Studie 1>	RCT, doppelblind, parallel	Jugendliche und Erwachsene, leichtes bis mittelschweres Asthma	<Gruppe 1> (n= 354) <Gruppe 2> (n= 347)	Run-in: 2 Wochen Behandlung: 6 Monate	Europa (Deutschland, Frankreich, Polen) 9/2003 – 12/2004	FEV1; Asthma-Symptome, Exazerbationen, gesundheitsbezogene Lebensqualität, unerwünschte Ereignisse

Tabelle 4-10: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	<Gruppe 1>	<Gruppe 2>	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in-Phase etc.</i>
<Studie 1>	xxx 250 µg, 1 Inhalation bid  + Placebo 2 Inhalationen bid	yyy 200 µg, 2 Inhalationen bid  + Placebo 1 Inhalation bid	Vorbehandlung: zzz 1000 µg pro Tag, 4 Wochen vor Studienbeginn  Bedarfsmedikation: aaa

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

Studie Gruppe	N	Alter (Jahre)	Geschlecht w/m (%)	<i>ggf. weitere Spalten mit Populationscharakteristika z. B. Dauer der Erkrankung, Schweregrad, weitere Basisdaten projektabhängig</i>
<Studie 1> <Gruppe 1> <Gruppe 2>				

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

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

Nicht zutreffend.

#### **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 4-12: 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			
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>

Begründen Sie für jede Studie die abschließende Einschätzung.

Nicht zutreffend.

#### 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 4-13: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

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

##### 4.3.1.3.1 <Endpunkt xxx> – RCT

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

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

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

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

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

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

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

Tabelle 4-14: 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 4-15: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
<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 4-16: Ergebnisse für <Endpunkt xxx> aus RCT mit dem zu bewertenden Arzneimittel

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

Nicht zutreffend.

*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.*

<Abbildung Meta-Analyse>

Abbildung 2: Meta-Analyse für <Endpunkt xxx> aus RCT; <zu bewertendes Arzneimittel> versus <Vergleichstherapie>

Nicht zutreffend.

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

#### **4.3.1.3.2 Subgruppenanalysen – RCT**

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

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

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

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

Nicht zutreffend.

#### **4.3.1.3.3 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien**

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

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

Nicht zutreffend.

#### **4.3.2 Weitere Unterlagen**

##### **4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien**

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

###### **4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche**

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

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

Nicht zutreffend.

###### **4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche**

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

Nicht zutreffend.

#### 4.3.2.1.3 Ergebnisse aus indirekten Vergleichen

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

Tabelle 4-17: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche

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

##### 4.3.2.1.3.1 <Endpunkt xxx> – indirekte Vergleiche aus RCT

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

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

Tabelle 4-18: 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 4-19: Operationalisierung von &lt;Endpunkt xxx&gt;

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 4-20: 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 4-21: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche

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

Nicht zutreffend.

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

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

*Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt, für den ein indirekter Vergleich vorgenommen wird, fortlaufend in einem eigenen Abschnitt dar.

#### **4.3.2.1.3.2 Subgruppenanalysen – indirekte Vergleiche aus RCT**

*Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen auf Basis indirekter Vergleiche aus RCT. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.*

Nicht zutreffend.

#### 4.3.2.2 Nicht randomisierte vergleichende Studien

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn nicht randomisierte vergleichende Studien als Nachweis für einen Zusatznutzen herangezogen werden sollen.

##### 4.3.2.2.1 Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien

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

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

Nicht zutreffend.

##### 4.3.2.2.2 Charakteristika der nicht randomisierten vergleichenden Studien

*Charakterisieren Sie nachfolgend die nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.*

*Beschreiben Sie die Verzerrungsaspekte der nicht randomisierten vergleichenden Studie auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Beschreibung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.*

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

Tabelle 4-22: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
			Patient	Behandelnde Personen		
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Studienebene.

Nicht zutreffend.

### 4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien

#### 4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien

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

Tabelle 4-23: 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 4-24: 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.

#### **4.3.2.3 Weitere Untersuchungen**

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn über die in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 genannten Studien hinausgehende Untersuchungen als Nachweis für einen Zusatznutzen herangezogen werden sollen.



**4.3.2.3.1 Ergebnis der Informationsbeschaffung – weitere Untersuchungen**

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung nach Untersuchungen, die nicht in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 aufgeführt sind. **Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie**

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

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

**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.**

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

**4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen****4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen**

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

Tabelle 4-25: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt. Ergebnisse nichtrandomisierter Studien, die keine kontrollierten Interventionsstudien sind,

*gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.*

Nicht zutreffend.

*Stellen Sie die Ergebnisse der weiteren Untersuchungen gemäß den jeweils gültigen Standards für die Berichterstattung dar. Begründen Sie dabei die Auswahl des Standards für die Berichterstattung. Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus weiteren Untersuchungen fortlaufend in einem eigenen Abschnitt dar.

#### **4.3.2.3.2 Subgruppenanalysen – weitere Untersuchungen**

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

Nicht zutreffend.

#### **4.3.2.4 Zusammenfassung der Ergebnisse aus weiteren Unterlagen**

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse aus weiteren Unterlagen (Abschnitte 4.3.2.1, 4.3.2.2 und 4.3.2.3) geben. Die Zusammenfassung soll Aussagen zu allen in diesen Abschnitten präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Meta-Analysen einschließlich Konfidenzintervallen dargestellt werden.

*Fassen Sie die Ergebnisse aus weiteren Unterlagen zusammen.*

Nicht zutreffend.

### **4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens**

#### **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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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.*

##### **4.4.2.1 Methoden**

###### **Hintergrund**

Cerliponase alfa (Brineura®) ist zur Behandlung eines seltenen Leidens indiziert und ist gemäß der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 als Arzneimittel für seltene Leiden („Orphan Drugs“) zugelassen [14]. Für Arzneimittel, die zur Behandlung eines seltenen Leidens zugelassen sind, gilt gemäß § 35a Sozialgesetzbuch (SGB) V Absatz 1 Satz 10 Halbsatz 1 der medizinische Zusatznutzen durch die Zulassung als belegt. Nachweise nach Satz 3 Nummer 2 und 3 müssen nicht vorgelegt werden [15]. Paragraph 5 Absatz 7 der VerfO bleibt hiervon unberührt; das Ausmaß des Zusatznutzens ist gemäß Paragraph 12 Nr. 1 Satz 2 der VerfO für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, nachzuweisen [16]. Diese Vorgaben sind auf Arzneimittel zur Behandlung seltener Leiden beschränkt, die den Umsatz von 50 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen [15].

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 von Modul 4 zu bearbeiten. Die Angaben zum Ausmaß des Zusatznutzens basieren auf der Zulassung und den der Zulassung zugrunde liegenden Studien [18].

### **Fragestellung**

Cerliponase alfa (Brineura®), eine humane, rekombinante Form der Tripeptidyl-Peptidase 1 (TPP1), ist zur Enzymersatztherapie (ERT) der neuronalen Ceroid-Lipofuszinose Typ 2 (CLN2) indiziert, einer seltenen, autosomal-rezessiv vererbten, lysosomalen Speicherkrankheit. CLN2, eine Form der Batten-Krankheit, wird durch Mutationen des CLN2-Gens verursacht, welches das lysosomale Enzym TPP1 kodiert. TPP1 spaltet Tripeptide vom aminoterminalen Ende kleiner Polypeptide und besitzt im sauren Milieu auch eine Endopeptidase-Aktivität [5]. Die Erkrankung ist durch einen Mangel der Enzymaktivität von TPP1 charakterisiert [5,9–11]. Dies führt zu einer Akkumulation von Speichermaterial in den Lysosomen und anschließend fortschreitender, neuronaler Degeneration [9,12]. TPP1 wird in verschiedenen Zelltypen exprimiert, neuronale Gewebe werden durch die Inaktivierung der TPP1 allerdings am meisten betroffen [3].

Die CLN2-Erkrankung ist gekennzeichnet durch Ataxie, Bewegungsstörungen (Myoklonus, Dystonie und Chorea), den Verlust motorischer und sprachlicher Fähigkeiten, fortschreitende Demenz und zunehmenden Sehverlust [2,3,5]. Krampfanfälle in Form tonisch-klonischer, partieller oder myoklonischer Anfälle oder Absencen [2–4] können behandlungsrefraktär werden [2]. Im Alter von etwa drei Jahren treten erste motorische und sprachliche Störungen auf und diese schreiten rasch voran bis zum nahezu kompletten Verlust dieser Fähigkeiten nach circa 2,5 Jahren. Die meisten Patienten werden im Verlauf der Erkrankung bettlägerig, erblinden und sind nicht mehr in der Lage, sich selbständig zu ernähren, da sie die Fähigkeit zu schlucken verlieren [3,6]. Die Patienten leiden in der späten Phase an Myoklonus und Schmerzen und sterben im Alter von 8-12 Jahren [5] [9,29].

Die in die klinische Studie 190-201 eingeschlossenen Patienten sowie die diesen zugeteilten Patienten der historischen Kontrollgruppe (190-901) befinden sich in einer hochprozedienten Phase der Erkrankung („rapid decline phase“; siehe Abbildung 3 und Abbildung 8), die hauptsächlich durch den Verlust motorischer Funktionen sowie des Sprachvermögens gekennzeichnet ist. Die motorischen und sprachlichen Fähigkeiten (gemessen mittels einer spezifischen Skala, der ML/HML-Skala; siehe Tabelle 4-29) gehen in der Progressionsphase in nahezu linearer Weise verloren (Abbildung 9); demgegenüber bestehen andere Ausprägungen der CLN2 wie Krampfanfälle oder unwillkürliche Bewegungen (Myoklonus) bis in das Endstadium fort. Die Motor- und Sprachfunktion repräsentieren die verlässlich messbaren Parameter der Krankheitsprogression und die entsprechenden Skalen des eingesetzten Messinstruments erfassen einen breiten Bereich des klinischen Krankheitsverlaufs. Daher wurde die 0-6-Punkte- ML/HML-Skala, welche die motorischen und sprachlichen Fähigkeiten erfasst, zur Bewertung der Krankheitsprogression für die Primäranalyse herangezogen.

Der klinischen Entwicklung von Cerliponase alfa lag die Hypothese zugrunde, dass intrazerebral verabreichte TPP1 die toxische Substratakkumulation und daraus folgend den

neuronalen Zelluntergang abschwächt. Dies wird durch die MRI-Resultate aus der klinischen Studie unterstützt. Ausgehend von dieser Hypothese kommt es zu einer Stabilisierung neuronaler Funktionen und bereits erworbene Fähigkeiten bleiben länger erhalten als bei Patienten unter heutigen Versorgungsstandards. In diesem Zusammenhang ist es von Bedeutung, dass die Neurodegeneration irreversibel ist und daher eine möglichst frühe Intervention anzustreben ist [30].

### ***Patientenpopulation***

Das vorliegende Dossier bezieht sich auf die Zulassung von Cerliponase alfa zur Behandlung der neuronalen Ceroid-Lipofuszinose (CLN2), auch bekannt als TPP1-Mangel.

### ***Intervention***

Cerliponase alfa (Brineura®) ist eine rekombinante Form der humanen TPP1 und wird in chinesischen Hamster-Ovarialzellen produziert. Brineura®-Lösung zur ICV-Injektion enthält 30 mg/ml Cerliponase alfa (150 mg/5 ml). Cerliponase alfa ist über einen implantierten Port (Reservoir und Katheter) intracerebroventrikulär (ICV) zu verabreichen, da es sich um ein großes Molekül handelt, welches die Blut-Hirn-Schranke nicht passiert. Der ICV-Port muss vor der ersten Infusion chirurgisch implantiert werden. In den Zulassungsstudien wurde Cerliponase alfa in einer Dosis von 300 mg alle zwei Wochen verabreicht (abgesehen von der Dosisfindungsphase der Studie 190-201).

### ***Vergleichstherapie***

Wie bereits oben dargestellt entfällt bei Arzneimitteln, die für seltene Leiden zugelassen sind, die Verpflichtung zum Nachweis des medizinischen Nutzens und Zusatznutzens im Verhältnis zur zweckmäßigen Vergleichstherapie, vorausgesetzt die festgelegte Umsatzgrenze (50 Millionen Euro in den vergangenen 12 Kalendermonaten) wird nicht überschritten [15]. Daher ist eine zweckmäßige Vergleichstherapie für Cerliponase alfa nicht zu benennen.

Zur Behandlung der CLN2 stehen derzeit weder innerhalb der Europäischen Union (EU) noch weltweit akzeptierte oder zugelassene Therapien zur Verfügung, die über rein unterstützende Maßnahmen hinausgehen. Zur symptomatischen oder palliativen Behandlung der Epilepsie (oft behandlungsresistent), des Myoklonus, der Spastizität oder der Schmerzen werden verschiedene Medikationen eingesetzt [3,10,31]. Andere Maßnahmen werden genutzt, um die Mobilität zu verbessern oder die Ernährung zu erleichtern bzw. das Risiko einer Aspiration zu vermindern (Gastrostomie). Psychotrope und Verhaltenstherapien können bei psychiatrischen Problemen und Verhaltensauffälligkeiten sinnvoll sein. Diese Behandlungsstrategien sind oft unbefriedigend und es liegen keine evidenzbasierten Therapieempfehlungen vor [30].

In Übereinstimmung mit der Europäischen Arzneimittel-Agentur (EMA) wurde aus ethischen Gründen auf eine Placebo-kontrollierte Studie verzichtet, da die Therapie die chirurgische Implantation eines ICV-Ports erfordert.

### **Endpunkte**

Entsprechend § 2 (3) der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) ist der Nutzen eines Arzneimittels der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität [32].

Zur Bewertung des Ausmaßes des Zusatznutzens von Cerliponase alfa (Brineura) werden die in den Zulassungsstudien erhobenen, patientenrelevanten Endpunkte herangezogen. Diese umfassen die Krankheitsprogression gemessen mittels einer krankheitsspezifischen Bewertungsskala (ML/HML), die Lebensqualität (QoL) und Nebenwirkungen.

### **Studientyp**

Zwei nicht kontrollierte Interventionsstudien mit Cerliponase alfa wurden initiiert: Eine Dosis-Eskalationsstudie der Phase 1/2 mit dem Ziel, die Sicherheit, Pharmakokinetik und Wirksamkeit zu untersuchen (Studie Nr. 190-201), und eine fortlaufende Langzeitstudie (Studie Nr. 190-202) im Anschluss an die Studie 190-201 (Extensionsstudie).

Bei der Studie 190-901 handelt es sich um eine andauernde, nicht interventionelle Studie zum natürlichen Verlauf der CLN2 (Patientenregister), um die Erkrankung zu beschreiben und die Krankheitsprogression bei unbehandelten CLN2-Patienten unter Einsatz eines krankheitsspezifischen Instruments (Hamburg Motor-Language [HML] scale) quantitativ zu erfassen.

Die Fragestellung wird auf Basis der Zulassungsstudie (und deren fortlaufender Folgestudie) sowie der Studie zum natürlichen Verlauf bei unbehandelten CLN2-Patienten (historische Kontrolle) bearbeitet. Die Bewertung des Behandlungseffekts bleibt auf diejenigen patientenrelevanten Endpunkte beschränkt, die sowohl in den klinischen Studien mit Cerliponase alfa als auch der historischen Kontrolle unter Einsatz vergleichbarer Messinstrumente erhoben wurden. Tabelle 4-26 gibt einen kurzen Überblick der Studien des pharmazeutischen Unternehmers.

Tabelle 4-26: Liste der Studien des pharmazeutischen Unternehmers – Weitere Untersuchungen

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer	Therapiearme
190-201	ja	ja	abgeschlossen	48 Wochen	Cerliponase alfa
190-202 <sup>1</sup>	ja	ja	laufend	Bis zu 239 Wochen	Cerliponase alfa
190-203 [33]	nein	ja	laufend	96 Wochen	Cerliponase alfa

<sup>1</sup> Da die Studie 190-202 nicht abgeschlossen ist und lediglich Zwischenanalysen vorliegen, werden die bislang vorliegenden Ergebnisse zur unterstützenden Information berichtet, beispielsweise um die Dauer des Effekts und die Sicherheit unter Langzeittherapie dazustellen.

Die Information in Tabelle 4-26 bildet den Studienstatus im Juni 2017 ab.

Tabelle 4-27: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – Weitere Untersuchungen

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
190-203	Die Studie ist nicht abgeschlossen und es handelt sich nicht um eine Zulassungsstudie.

### Validität und Patientenrelevanz der Endpunkte

In die pivotale Studie wurden Patienten in einer frühen oder mittleren Phase der Krankheitsprogression eingeschlossen, um während der Studiendauer die Phase der rasch fortschreitenden Krankheitsprogression zu erfassen, und eine vergleichbare Patientenpopulation zur historischen Kontrolle zu gewährleisten. Obwohl es sich bei der CLN2 um eine sehr schwere, rasch voranschreitende und letztlich tödliche Erkrankung handelt, sind die Behandlungsdauer von mindestens 48 Wochen in der Studie 190-201 und die Dauer der Nachbeobachtung bis zum Datenschnitt der Studie 190-202 zu kurz, um einen messbaren Effekt auf das Überleben erwarten zu können. Bislang wurden keine Todesfälle berichtet [34]. Derzeit kann der Einfluss der Behandlung auf die Mortalität noch nicht erfasst werden, dies wird jedoch in den folgenden Jahren möglich sein.

Zur Bewertung des Ausmaßes des Zusatznutzens von Cerliponase alfa werden die folgenden patientenrelevanten Endpunkte herangezogen:

- Krankheitsaktivität (Morbidität) gemessen mittels eines CLN2-spezifischen Instruments: ML-Skala (motorische/sprachliche Fähigkeiten)  
Als Sensitivitätsanalysen wurden zusätzlich zur ML-Skala folgende Summen-Scores ermittelt: MLV-Skala (motorische/sprachliche Fähigkeiten plus Sehvermögen [vision]),

MLVS-Skala (motorische/sprachliche Fähigkeiten/Sehvermögen plus Krampfaktivität [seizures]); motorische und Sprachdomäne getrennt

- Lebensqualität (QoL): generisches Instrument PedsQL™ (Module Elternbericht für Kleinkinder und familiäre Belastung) sowie ein krankheitsspezifisches (CLN2-spezifisches) QoL-Messinstrument
- Unerwünschte Ereignisse (UEs; behandlungs- und krankheitsbedingte Morbidität)

### **Krankheitsaktivität auf krankheitsspezifischen Bewertungsskalen**

Zur klinischen Bewertung der Schwere der CLN2 wurden in der Vergangenheit zwei ähnliche Skalen eingesetzt, die so genannte Hamburg-Skala [5,35] sowie eine Skala, die das „Weill Cornell Medical College“ (WCMC-Skala) entwickelt hat [6,36]. Beide Instrumente erfassen krankheitsspezifische, klinische Parameter auf einer Skala von 0-12 Punkten. Die Instrumente sind geeignet, um die Fähigkeiten und den Krankheitszustand von Kindern ab zwei Jahren über den gesamten, weiteren Krankheitsverlauf quantitativ zu bestimmen. Jede der Skalen besteht aus vier Domänen: Hamburg-Skala (motorische und Sprachfähigkeit, Sehen und Krampfanfälle) und WCMC-Skala (Gehen, Sprache, Myoklonus und Ernährung). Jede Domäne beider Instrumente umfasst eine Skala von 0-3 Punkten, wobei 3 Punkte die altersgemäß beste Funktion und 0 Punkte den schlechtesten Zustand bzw. einen nahezu vollständigen Funktionsverlust beschreiben (Tabelle 4-28). Die motorische und Sprachdomäne sind von höchster klinischer Relevanz und dementsprechend Domänen beider Instrumente, die von zwei Arbeitsgruppen unabhängig voneinander entwickelt wurden. Die Graduierungen zur Beschreibung der Mobilität (motorische Funktion) und Sprache sind bei beiden Skalen ähnlich.

Tabelle 4-28: Hamburg und WCMC-Bewertungsskalen zur quantitativen Erfassung der Krankheitsaktivität

<b>Hamburg Skala [5]</b>			<b>WCMC Skala [6]</b>		
Motorische Fähigkeiten	3	Geht normal	Gang	3	Normal
	2	Häufige Stürze, offensichtliche Ungeschicklichkeit		2	Abweichend (spastisch oder bradykinetisch oder ataxisch), aber in der Lage, selbstständig zu gehen
	1	Kein Gehen ohne Hilfe oder nur Krabbeln		1	Abweichend (spastisch oder bradykinetisch oder ataxisch), benötigt Hilfe
	0	Immobil, meist bettlägerig		0	Nicht gehfähig
Sprache	3	Normal	Sprache	3	Normale Sprache
	2	Erkennbar von der Norm abweichend		2	Von der Norm abweichende Sprache mit ungewöhnlicher Artikulation und vermindertem Wortschatz



Hamburg Skala [5]			WCMC Skala [6]		
	1	Schwer verständlich		1	Kaum verständliche Sprache mit schwerer Dysarthrie und nur sehr wenigen verständlichen Worten
	0	Unverständlich oder keine Sprache		0	Nicht verständliche Worte oder keine Sprache
Sehvermögen	3	Erkennt ein gewünschtes Objekt, greift nach diesem	Myoklonus	3	Kein Myoklonus, Chorea/Tremor/Athetose und Dorsalextension der Zehen („upgoing toes“)
	2	Unkoordiniertes Greifen nach Objekten		2	Eines der Folgenden: Myoklonus, Chorea/Tremor/Athetose oder Dorsalextension der Zehen („upgoing toes“)
	1	Reagiert auf Licht		1	Zwei der Folgenden: Myoklonus, Chorea/Tremor/Athetose oder Dorsalextension der Zehen („upgoing toes“)
	0	Keine Reaktion auf visuelle Reize		0	Myoklonus, Chorea/Tremor/Athetose und Dorsalextension der Zehen („upgoing toes“)
Krampfanfälle (Grand mal)	3	Keine Krampfanfälle in drei Monaten	Ernährung	3	Keine Schluckstörung
	2	Ein bis zwei Krampfanfälle in drei Monaten		2	Milde Schluckstörung
	1	Ein Krampfanfall pro Monat		1	Mittelschwere Schluckstörung
	0	Mehr als ein Krampfanfall pro Monat		0	Gastrostomie-Schlauch-abhängig

Der primäre Endpunkt der Studie 190-201/202 ist der Summenscore aus der motorischen (Mobilitäts)- und Sprachdomäne (0-6 Punkte). Beide Domänen der ursprünglichen Hamburg-Skala (HML) wurden für die Studie 190-201 geringfügig verändert, um einerseits objektive Ankerpunkte zu erhalten und andererseits die Abgrenzung zwischen den Kategorien zu präzisieren. Diese Änderungen wurden während eines Treffens von CLN2-Experten und Vertretern der ersten beiden Studienzentren festgelegt. Hierbei wurde sichergestellt, dass die beiden Messinstrumente miteinander vergleichbare Ergebnisse liefern [24].

Sowohl die ursprüngliche HML-Skala [5] als auch die geringfügig modifizierte ML-Skala umfassen zwei Domänen, die durch den Prüfarzt erhoben werden. Die motorischen und sprachlichen Fähigkeiten werden auf einer 4-Punkt-Skala bewertet (von 3-0 Punkte), wobei die einzelnen Abstufungen innerhalb beider Domänen altersgerecht normale Fähigkeiten (3 Punkte), bis hin zum vollständigen Funktionsverlust (0 Punkte) beschreiben [24]. Tabelle 4-29 zeigt einen Vergleich der HML-Skala und der geringfügig modifizierten ML-Skala. Abgesehen von der motorischen Domäne und der Sprachdomäne blieben die übrigen Domänen unverändert.

Tabelle 4-29: Vergleich der HML Skala mit der ML Skala

HML-Skala [5]			ML-Skala [37]
Motorische Fähigkeiten (Hamburg)	3	Geht normal <sup>1</sup>	Weitgehend normaler Gang. Keine auffällige Ataxie, keine krankheitsbedingten Stürze.
	2	Häufige Stürze, offensichtliche Ungeschicklichkeit	Selbständiges Gehen, gekennzeichnet durch die Fähigkeit, $\geq 10$ Schritte ohne Hilfe zu gehen. Offensichtliche Unsicherheit und gelegentliche Stürze
	1	Kein Gehen ohne Hilfe oder nur Krabbeln	Benötigt externe Hilfe zum Gehen oder kann nur Kriechen
	0	Immobil, meist bettlägerig	Kann nicht länger gehen oder kriechen
Sprachvermögen(Hamburg)	3	Normal (individuell beste Leistung) <sup>2</sup>	Augenscheinlich normale Sprache. Verständlich und weitgehend altersadäquat. Noch keine Verschlechterung festzustellen.
	2	Erkennbar von der Norm abweichend	Sprache hat sich verschlechtert, von der Norm abweichend (schlechter als die individuell beste Leistung) <sup>3</sup>
	1	Schwer verständlich	Schwer verständlich. Nur wenige verständliche Worte.
	0	Unverständlich oder keine Sprache	Keine verständlichen Worte oder Vokalisation
<p><sup>1</sup> Bei einigen Kindern war die motorische Entwicklung nie altersadäquat.</p> <p><sup>2</sup> Bei einigen Kindern war die Sprachentwicklung nie altersadäquat. In diesen Fällen wurde die beste Leistung, die jemals erreicht wurde, als Ausgangspunkt mit Drei bewertet. Wenn Sprache erkennbar schlechter wurde, wurde ein Score von Zwei gegeben.</p> <p><sup>3</sup> Einige verständliche Worte; ist in der Lage, kurze Sätze zu bilden, um Begriffe, Wünsche oder Bedürfnisse zu artikulieren. Der Score beschreibt eine deutliche Verschlechterung bezogen auf die früheren Fähigkeiten (d.h. des bislang besten Ergebnisses des Kindes). Ob sich der Zustand des Kindes verschlechtert hat, kann nur durch Befragung der Eltern geklärt werden.</p>			

Phänotypisch ist bei der CLN2 hauptsächlich das Zentralnervensystem (ZNS) betroffen [38]. Obwohl in der Studie 190-201 alle vier Domänen der Hamburg- und WCMC-Skala erhoben wurden, werden diejenigen Domänen, die beiden Skalen gemeinsam sind (motorische Funktion [HML]/Gang [WCMC] sowie Sprache), als besonders relevant für Patienten in einem frühen oder mäßig fortgeschrittenem Krankheitsstadium erachtet [37]. Da die motorischen und sprachlichen Fähigkeiten während der Phase der schnellen Krankheitsprogression linear

abnehmen und beide Domänen alle Krankheitsstadien von altersgerecht normal bis zum kompletten Verlust der Funktionen abdecken, wurde der 0-6-Punkt-Summenscore der ML-Skala zur Primäranalyse eingesetzt.

Krampfanfälle treten bereits in einem frühen Stadium der Erkrankung auf, jedoch auch während der gesamten Krankheit in unterschiedlicher Häufigkeit und Schwere. Die Krampfneigung der Patienten unterliegt deutlichen Schwankungen (siehe Abbildung 7) und wird durch Begleitmedikation nur zu einem gewissen Grad moduliert. Da die Häufigkeit und Schwere der Krampfanfälle sowohl zwischen den Patienten als auch intra-individuell stark variiert, wird dieser Endpunkt im Gegensatz zur Mobilitäts- und Sprachkomponente als weniger sensitiv eingestuft. Darüber hinaus wird die Krampfneigung der Patienten durch die anti-epileptische Therapie beeinflusst.

Das Sehvermögen wird erst in einem späten Krankheitsstadium verloren [39]. In der Studie 190-901 traten Sehbehinderungen später auf als die motorischen und Sprachstörungen (Abbildung 5, Abbildung 6) [39]. Chang et al. (2011) beschreibt die ophthalmologische Symptomatik wie folgt: “A gradual decline in visual ability is observed in children with CLN2 disease from 4 years of age, leading to complete blindness often within ~3 years. However, in some patients, blindness may not become apparent before the age of 10 years” [3]. Zur Behandlung der ophthalmologischen Manifestationen stehen derzeit keine Therapien zur Verfügung [40].

#### *Patientenrelevanz*

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patienten-relevante therapeutische Effekt insbesondere hinsichtlich der *Verbesserung des Gesundheitszustandes*, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ [32].

Die Krankheitsprogression wurde in den Studien 190-201/202 und 190-901 anhand der Änderung auf der ML/HML-Skala über 48 Wochen erfasst (Score von 6 [weitgehend im Normbereich] bis 0 [vollständiger Funktionsverlust]). Als Schwellenwerte für das Ansprechen gemessen mittels Summenscore wurden bei der Primäranalyse <2 Punkte (Ansprechen auf die Therapie) bzw. ≥2 Punkte (Fortschreiten der Erkrankung) festgelegt.

Jede Zunahme oder Abnahme auf der ML/HML-Skala um einen Punkt entspricht einer leicht erkennbaren und deutlichen Veränderung der motorischen und sprachlichen Fähigkeiten des Patienten. Zum Beispiel bedeutet ein Abfall des motorischen Scores von zwei auf einen Punkt, dass ein Kind, das sich unabhängig allerdings ataxisch und unbeholfen fortbewegen kann, weitgehend gehunfähig wird oder sich nur noch mit Hilfen fortbewegen kann. Ein Abfall der Sprachkomponente von zwei auf einen Punkt bedeutet, dass ein Kind, dessen Sprache zu Beginn erkennbar von der Norm abweicht, nur noch zu wenigen verständlichen Worten in der Lage und schwer verständlich ist, d. h. eine Kommunikation ist fast nicht mehr möglich. Eine Änderung um zwei Punkte des Summenscores entspricht dabei entweder einer Änderung um zwei Punkte auf einer der Domänen oder um einen Punkt auf jeder der beiden Domänen.

Jede Abnahme des ML/HML-Scores um zwei Punkte kann als „Meilenstein“ der Krankheitsprogression angesehen werden, die eine Teilnahme am normalen Leben und übliche Alltagsaktivitäten zunehmend schwierig macht. Der Verlust entscheidender Entwicklungsschritte (abgebildet durch die vier Scores von 0-3 der Einzeldomänen) wie die Fähigkeit, mit der Familie zu kommunizieren oder sich selbständig fortzubewegen, ist klinisch relevant und für die Patienten von großer Bedeutung. Schätzungen des kleinsten, für die Patienten spürbaren Unterschieds („minimal important difference“, MID) mithilfe statistischer, verteilungsbasierter Methoden (siehe Kapitel 4.4.2.3.1.2) unterstützen dies nachdrücklich. Nach diesen Schätzungen entspricht eine Änderung um 0,6 Punkte auf der ML/HML-Skala der MID. Eine Änderung um zwei Punkte auf der ML/HML-Skala stellt daher eine deutlich größere Veränderung dar als die auf Grundlage einer statistischen Verteilungsmethode geschätzte MID. Dies rechtfertigt die Änderung um zwei Punkte als patientenrelevantes Erfolgskriterium. Da die mittlere Verschlechterung von unbehandelten Patienten in einem Jahr bei ca. 2 Punkten liegt, und in der Nachverfolgung ein Zeitraum von mindestens 96 Wochen beobachtet wurde, kann die klinische Studie 190-201/202 zu verlässlichen Aussagen zum Krankheitsverlauf von behandelten Patienten führen.

#### *Validierung des Messinstruments*

Die Hamburg-Skala, bestehend aus vier Domänen (HML), wurde erstmals in einer Längsschnittstudie von Steinfeld et al. (2002) bei Kindern mit CLN2 und Wohnsitz in Deutschland oder der Schweiz beschrieben. Die Einzel-Scores zu den motorischen, visuellen und sprachlichen Fähigkeiten wurden zu einem Gesamtscore („Total Disability Score“) zusammengefasst (maximaler Score von 9). Die Studie von Steinfeld hatte zum Ziel, den Krankheitsverlauf quantitativ zu beschreiben [5]. Diese und andere Studien zum natürlichen Verlauf der CLN2 (WCMC und Hamburg) belegen, dass die Bewertungsskala sensitiv gegenüber einer Krankheitsprogression ist und den vorhersagbaren Krankheitsverlauf abbildet.

Im Auftrag von Biomarin wurde ein „Clinician-Reported Outcome Evidence Dossier“ erstellt, in welchem die Performance und Validität der ML-Skala sowie deren Einzeldomänen detailliert untersucht wurden [24].

In einer Brückenstudie wurden Videobänder analysiert, die während der Erhebungen der ML-Skala in den Studien 190-201/202 aufgezeichnet worden waren. Einer der Entwickler der HML-Skala, der nicht an der Studie 190-201/202 beteiligt war, führte einen unabhängigen Review mit der Fragestellung durch, ob die motorische und Sprachdomäne der HML-Skala (Studie zum natürlichen Verlauf) und die in der Studie 190-201/202 eingesetzte, modifizierte ML-Skala hinreichend ähnlich sind. Ein Angestellter von Biomarin, der keinen Zugang hatte und nicht an den Studien 190-201/202 beteiligt war, überprüfte die Videos bevor diese dem unabhängigen Gutachter übergeben wurden. Zur Überprüfung der Übereinstimmung zwischen den Bewertungen durch den unabhängigen Gutachter unter Einsatz der HML-Skala (motorisch, Sprache und motorische und Sprach-Scores kombiniert) und denjenigen der Prüfarzte der Studie 190-201/202 auf Basis der ML-Skala wurde Cohens Kappa („weighted kappa“,  $K_w$ ) als statistisches Maß eingesetzt [41]. Die Ergebnisse zeigten zu allen Untersuchungszeitpunkten (Baseline, Woche 25, 48 und 72 [Woche 25 der Studie 190-202]) eine beachtliche bis perfekte

Übereinstimmung ( $0,67 \leq K\omega \leq 1,00$ ) zwischen den beiden motorischen Skalen (Kappa-Bereich), eine ausreichende bis beachtliche Übereinstimmung zwischen den Sprachskalen ( $0,34 \leq K\omega \leq 0,62$ ) und eine beachtliche Übereinstimmung hinsichtlich der kombinierten motorischen und Sprach-Scores ( $0,67 \leq K\omega \leq 0,79$ ). Darüber hinaus ergaben die Analysen der mittleren Steigung („slope“) der HML- und ML-Skala, berechnet über den Zeitraum zwischen Baseline und Woche 48, nahezu identische Punktschätzer und Variabilität. Dies traf auf die Steigungen der motorischen Skala ( $0,14 \pm 0,25$  [HML] vs.  $0,17 \pm 0,33$  [ML]), der Sprachskala ( $0,00 \pm 0,00$  [HML] vs.  $-0,02 \pm 0,575$  [ML]) und der aus den Werten der motorischen und Sprach-Subskala zusammengesetzten 0-6-Skala ( $0,05 \pm 0,41$  [HML] vs.  $0,17 \pm 0,52$  [ML]). Die Ergebnisse stellen sicher, dass die HML- und ML-Skalen hinreichend ähnlich sind und eine Interpretation der Ergebnisse der Studien 190-201 und 190-202 erlauben [24].

Zusätzlich wurden in einer formalen Inter-Rater-Studie insgesamt 36 Videoaufzeichnungen zu vier Zeitpunkten (Baseline, Woche 25, 48 und 72 [Woche 25 der Studie 190-202]) bei 12 der in Hamburg rekrutierten Patienten der Studie 190-201/202 analysiert. Ein Arzt, der die Prüfarzte der Studie hinsichtlich der Anwendung der ML-Skala ausgebildet hatte, führte diese Studie durch. Die Bewertungen des Arztes wurden mit denjenigen der Prüfarzte verglichen. Zur statistischen Analyse der Übereinstimmung zwischen den Bewertungen des Arztes und der Prüfarzte wurde Cohens Kappa („weighted kappa“,  $K\omega$ ) eingesetzt [41] und die Klassifizierung von Landis und Koch (1977) [42], um die Inter-Rater-Reliabilität zu charakterisieren. Hinsichtlich der motorischen Scores war die Übereinstimmung in Woche 25 ( $n=11$ ), Woche 48 ( $n=10$ ) und Woche 72 ( $n=3$ ) bei einem  $K\omega$  von 1,00 perfekt. Zu Baseline bewertete der Arzt zwei der motorischen Skalenwerte einen Punkt niedriger als die Prüfarzte ( $K\omega = 0,76$ ;  $n=12$ ), dennoch war die Übereinstimmung beachtlich. Bei einem  $K\omega$  von 0,82 war die Übereinstimmung der Bewertungen der Sprach-Scores aller Videos nahezu perfekt. Die  $K\omega$ -Schätzungen schwankten über alle Erhebungszeitpunkte zwischen 0,67 in Woche 48 und 0,93 zu Baseline. Darüber hinaus stimmten die Bewertungen der kombinierten ML-Scores (motorische Funktion plus Sprache; 0-6) zu allen Untersuchungszeitpunkten und -Intervallen durch den Arzt nahezu perfekt mit denjenigen der Prüfarzte überein ( $0,89 \leq K\omega \leq 0,93$ ). Dieses Ergebnis zeigt die hohe Inter-Rater-Reliabilität hinsichtlich der ML-Scores [24].

Die psychometrischen Eigenschaften der ML-Skala einschließlich des Gesamt-Scores (0-6) und der motorischen und Sprach-Scores getrennt wurden auf Basis der Daten der klinischen Studie 190-201 untersucht. Zur Analyse der Konstruktvalidität wurde das Muster und Ausmaß des Zusammenhangs zwischen der motorischen und Sprachkomponente sowie der daraus zusammengesetzten ML-Scores zu Baseline mit anderen, ähnlichen Endpunkten mittels Spearman-Korrelation verglichen. Die motorischen Scores wurden mit drei anderen Domänen (körperliche, emotionale und soziale Funktion) sowie dem Gesamtscore des PedsQL™-Moduls „Familiäre Belastung“ verglichen [43]. Zusätzlich wurden die Ergebnisse zu Alltagsaktivitäten aus dem CLN2-spezifischen Fragebogen zur Lebensqualität (CLN2 QoL) untersucht, um die Konstruktvalidität der motorischen Funktionsskala näher zu charakterisieren. Der Sprach-Score wurde denjenigen Domänen des PedsQL™-Moduls „Familiäre Belastung“ verglichen, die die soziale und kognitive Funktion sowie Kommunikation betreffen. Diese stellen ein Maß dafür dar, welchen Einfluss das Sprachvermögen des Kindes auf die elterliche Rollenfunktion hat.

Schließlich wurde der kombinierte ML-Score (motorische plus Sprachfunktion) mit folgenden Endpunkten verglichen: PedsQL™-Gesamt-Score, PedsQL™-Gesamt-Score Modul „Familiäre Belastung“ und CLN2 QoL-Score „Alltagsaktivitäten“ (zur Operationalisierung siehe Tabelle 4-53 und Tabelle 4-55) [24]. Bei allen Korrelationsanalysen wurde mithilfe des Spearman-Korrelationskoeffizienten untersucht, ob ein schwacher ( $r < 0,30$ ), mittlerer ( $0,30 \leq r < 0,60$ ) oder starker ( $r \geq 0,60$ ) Zusammenhang besteht.

Die Korrelationen zwischen der motorischen Domäne und dem PedsQL™ Kernmodul, dem Modul „Familiäre Belastung“ sowie dem CLN2-QoL-Fragebogen waren zu Baseline grundsätzlich moderat bis hoch ( $0,40 \leq r \leq 0,64$ ), wobei die Korrelation mit dem Endpunkt „Körperliche Funktion“ (PedsQL™) am höchsten war ( $n=23$  außer CLN2 QoL-Alltagsaktivitäten:  $n=22$ ). Auch die Korrelationen zwischen dem Baseline-Wert zum Sprachvermögen und dem PedsQL™-Modul „Familiäre Belastung“ und dem CLN2-QoL-Fragebogen waren moderat bis hoch ( $0,40 \leq r \leq 0,64$ ); der stärkste Zusammenhang bestand mit der Domäne „Kommunikation“ des PedsQL™ „Familiäre Belastung“ ( $n=23$ ). Die Analyse ergab moderate Korrelationen ( $0,35 \leq r \leq 0,51$ ) zwischen dem Gesamt-Score der modifizierten ML-Skala (0-6) und dem PedsQL™ Gesamt, PedsQL™ „Familiäre Belastung“ Gesamt und dem CLN2-QoL-Score „Alltagsaktivitäten“ [24].

Die Responsiveness der kombinierten motorischen und sprachlichen Domäne der ML-Skala sowie der Einzel-Scores wurde anhand des Korrelationskoeffizienten nach Spearman untersucht. Die Änderung der ML-Scores zwischen Baseline und Woche 48 (Studie 190-201 Studienende/vorzeitiger Abbruch) wurden mit den Änderungen des CLN2-QoL-Fragebogens „Alltagsaktivitäten“ zwischen Baseline und der jeweiligen Visite verglichen. Die Korrelationsanalyse zwischen der Änderung der motorischen Domäne bzw. des Gesamt-Scores und dem CLN2-QoL-Score „Alltagsaktivitäten“ zeigte eine moderate Responsiveness ( $r=0,37$  bzw.  $0,32$ ,  $n=22$ ). Die Korrelation zwischen der Änderung des PedsQL™ „Soziale Funktion“ und dem Sprachvermögen, gemessen auf der modifizierten Hamburg-Skala, betrug  $0,29$ ; entsprechend war die Korrelation mäßig ausgeprägt. Die ML-Skala änderte sich im Verlauf der Studie 190-201 nur wenig, sodass die Chance, irgendeine hohe Korrelation zu entdecken, gering war.

### *Operationalisierung*

Die Operationalisierung des Endpunkts ist in Tabelle 4-39 beschrieben. Die modifizierte ML-Skala wurde in den Studien 190-201/202 eingesetzt (in diesem Dokument durchgängig abgekürzt als „ML-Skala“), während die Studie zum natürlichen Verlauf (historische Kontrolle) die von Steinfeld et al. [5] beschriebene Original-Hamburg-Skala („HML-Skala“) verwendete. Die Unterschiede zwischen der ML- und HML-Skala werden als minimal eingeschätzt und die geringfügigen Änderungen der ML-Skala dienen ausschließlich dazu, aus Gründen der Klarheit objektive Ankerpunkte zu liefern. In der Studie 190-201/202 wurden folgende Endpunkte der modifizierten Skalen erhoben: 0-6-Punkt ML-Skala, 9-Punkt CLN2-Skala (MLV: motorisch-Sprache-visuell) und 12-Punkt CLN2-Gesamt-Score (motorisch-Sprache-visuell-Krampfanfälle). Die Studie 190-202 ist eine Erweiterungsstudie von 190-201 und die Operationalisierung des Endpunkts war in beiden Studien identisch.

*Schlussfolgerung:*

Zwei unabhängige Arbeitsgruppen bestätigen, dass Beeinträchtigungen der motorischen und sprachlichen Fähigkeiten bedeutende Krankheitsmanifestationen der CLN2 sind [5,6]: Sowohl die Domäne „Gang“ des WCMC als auch „motorische Funktion“ der Hamburg-Skala erfassen die Mobilität und Koordination der Patienten und die beiden Skalen gemeinsame Domäne „Sprache“ den Verlust des Sprachvermögens. Die Ergebnisse der Studie zum natürlichen Verlauf zeigen, dass die HML-Skala geeignet ist, zwischen verschiedenen Krankheitsstadien zu unterscheiden und den fortschreitenden Krankheitsverlauf abzubilden. Der Summen-Score, zusammengesetzt aus den Domänen motorische Funktion/Gang und Sprache, wurde als primärer Endpunkt der Studien 190-201/202 gewählt, da sich beide Domänen mit Fortschreiten der Erkrankung in sehr ähnlicher Weise (simultan) verschlechtern: ab einem Alter von etwa drei Jahren nehmen beide Fähigkeiten rasch ab und im Alter von circa fünf Jahren sind die Patienten nicht mehr in der Lage, sich fortzubewegen oder zu sprechen [5]. Beide Domänen werden für die frühe Progression der CLN2 als höchst relevant eingestuft. Da die motorische und Sprachdomäne Wendepunkte der Krankheitsprogression von altersgerecht normal bis zu nur noch minimalen oder völligem Verlust der Funktionen abbilden, ist die ML-Skala ein Messinstrument, das sowohl für die betroffenen Kinder als auch deren Eltern/Betreuer von Bedeutung ist.

Die Analysen der ML/HML-Skala im Auftrag von Biomarin durch unabhängige Gutachter zeigen eine hohe Inter-Rater-Reliabilität (Reanalyse der Videoaufzeichnungen verglichen mit den Bewertungen der Prüferärzte,  $0,89 \leq K\omega \leq 0,93$ ) und eine beträchtliche Übereinstimmung zwischen der unabhängigen Bewertung der Videoaufzeichnungen der Studie 190-201/202 unter Einsatz der HML-Skala bzw. der ML-Skala durch die Prüferärzte ( $0,67 \leq K\omega \leq 0,79$ ); dies bestätigt die Slope-Analyse der Brückenstudie, deren Ziel war, den Verlauf (Baseline bis Woche 48) unter Einsatz der HML-Skala verglichen mit der ML-Skala zu untersuchen. Die Konstruktvalidität, d. h. der Zusammenhang zwischen dem ML-Score zu Baseline und patientenberichteten Endpunkten, war moderat ( $0,35 \leq r \leq 0,51$ ), ebenso wie die Korrelation zwischen dem Verlauf des PedsQL™ „Soziale Funktion“ und dem Sprachvermögen von Baseline bis zu Woche 48 in der Studie 190-201 ( $r=0,29$ ). Zusammenfassend stützen die Analysen die Schlussfolgerung, dass das Instrument die Krankheitsprogression wie beabsichtigt erfasst.

***MRI-Parameter (nicht eingeschlossen)***

Im Rahmen eines „Scientific Advice“ betonte der Ausschuss für Humanarzneimittel (CHMP) die Bedeutung objektiver Parameter aus Gründen der Plausibilität und die besondere Bedeutung von MRI-Daten als potenzielle Biomarker der Krankheitsaktivität [44].

*Patientenrelevanz*

Da es sich um einen nicht validierten Surrogat-Endpunkt handelt, ist die Patientenrelevanz von MRI-Daten bei CLN2 noch unklar.

### *Validität*

Die Ergebnisse der Studie 190-201 zeigten eine geringe bis moderate Korrelation zwischen den MRI- und klinischen Befunden (ML-Skala). Das Gesamtvolumen der kortikalen, grauen Substanz, entweder ausgedrückt als absolute Änderung oder relativ zum gesamten Hirnvolumen (WBV), zeigte keinen eindeutigen Zusammenhang mit dem Verlauf der ML-Skalenwerte ( $r^2 = 0,20$  bzw.  $r^2 = 0,48$ ). Bei Patienten, deren ML-Score sich nicht verändert hatte, war der Volumenverlust vergleichbar mit Patienten, deren Wert sich um 1-2 Punkte verschlechtert hatte. Moderate Korrelationen bestanden zwischen der Abnahme des CSF-Volumens ( $r^2 = 0,61$ ) bzw. des WBV ( $r^2 = 0,52$ , beides absolute Änderung) und dem ML-Score [26]. MRI-Daten unbehandelter Patienten mit CLN2 werden in dem Zentrum in Hamburg erst seit 2009 erhoben [45] und es liegen keine angemessenen Längsschnittdaten zum natürlichen Verlauf der Erkrankung vor.

Insgesamt scheint die Korrelation zwischen dem Volumenverlust und dem klinischen Verlauf über diesen 48-Wochen-Zeitraum nur gering zu sein, dennoch sind weitere Langzeiterhebungen sowie Daten während der frühen Krankheitsprogression für ein besseres Verständnis dieser Korrelation erforderlich. Darüber hinaus sind weitere Untersuchungen wichtig, um zu überprüfen, ob Remodellierungseffekte die Korrelation zwischen klinischen und MRI-Daten zu frühen Zeitpunkten möglicherweise verzerren (z. B. scheinbare Volumenreduktion durch Abbau lysosomalen Speichermaterials).

Die MRI ist bei Patienten mit CLN2 nicht als Surrogatendpunkt validiert und wird daher nicht als patientenrelevanter Endpunkt eingeschlossen.

### *Operationalisierung*

Nicht zutreffend.

Schlussfolgerung: Die MRI-Ergebnisse werden bei der Bewertung des Zusatznutzens von Cerliponase alfa nicht berücksichtigt, da die Daten nicht ausreichen, um die MRI als Surrogatendpunkt bei CLN2-Patienten zu validieren, und keine ausreichenden Längsschnittdaten zum natürlichen Verlauf vorliegen.

### *Lebensqualität (QoL)*

Bei seltenen Erkrankungen stellt die Wahl der Endpunkte für klinische Studien oftmals eine Herausforderung dar. Zur Erfassung der Lebensqualität sollten stets Instrumente eingesetzt werden, die für die jeweilige, zu behandelnde Indikation validiert sind. Die Europäische Arzneimittelagentur (EMA) bestätigt in der "Guideline on clinical trials in small populations", dass die Patientenpopulationen zu klein sein können, um neben der Endpunktvalidierung auch den Behandlungseffekt zu untersuchen [46]. Eine Projektgruppe der ISPOR („International Society for Pharmacoeconomics and Outcomes Research“) hat im Rahmen eines Reviews verschiedene, selbstberichtete Messinstrumente untersucht, die bei Kindern und Jugendlichen eingesetzt werden. Bei Kindern unter fünf Jahren gibt es keine überzeugende Evidenz, die die Reliabilität und Validität von Messinstrumenten bei dieser Altersgruppe belegt. Dennoch liegen Erhebungsbögen vor, die von Kindern unter fünf Jahren selbst zu beantworten sind. In manchen



Situationen ist es sogar unmöglich, die Lage aus Sicht des Kindes zu erfassen z. B., wenn die Zielpopulation zu jung oder kognitiv beeinträchtigt ist, um Gesundheitsfragen zu lesen, zu verstehen oder zu beantworten. Daher ist man bei sehr jungen Patienten auf klinische Parameter oder die Beobachtungen der Eltern oder anderer Erwachsener angewiesen, um den Gesundheitszustand zu erfassen. Nach der „Food and Drug Administration“ (FDA) sollten bei pädiatrischen Patienten, die nicht in der Lage sind, selbst Angaben zu ihrem Gesundheitszustand zu machen, eher direkte Beobachtungen genutzt werden als Proxy-Instrumente, die auf subjektiven Erfahrungen des Kindes basieren [47].

Zur Messung der Lebensqualität wurden in den Studien 190-201/202 drei verschiedene Fragebögen eingesetzt: PedsQL™ „Measurement Model for Pediatric Quality of Life Inventory“ (PedsQL™), ein von Biomarin entwickelter, CLN2-spezifischer Fragebogen und der „EuroQol Health Status“ EQ-5D-5L (Letzterer nur in der Studie 190-202).

#### *Patientenrelevanz*

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer *Verbesserung der Lebensqualität*“ [32].

Eine Online-Umfrage unter 24 CLN2-Experten einschließlich Patientenvertreter hatte zum Ziel, Informationen über das Krankheitsmanagement zu erhalten. Die Lebensqualität (sowohl Kinder als auch Familie) stellte eine der vier Hauptsäulen des Managements dar. Therapieziele waren:

- Möglichst lange Teilnahme an der Schule sowie gesellschaftlichen und üblichen Aktivitäten zuhause,
- Mobilität und Unabhängigkeit erhalten,
- Kommunikation erhalten,
- Wohlbefinden erhalten [40].

#### *PedsQL™*

Der PedsQL™ wurde entwickelt, um die gesundheitsbezogene QoL bei Kindern und Jugendlichen zu erfassen. Es liegen generische und krankheitsspezifische Fragebögen vor [48]. In der Studie 190-201/202 wurde der generische Fragebogen des PedsQL™ eingesetzt, da kein krankheitsspezifisches Modul existiert. Das generische Kernmodul (Selbst- oder Elternbericht) liegt für verschiedene Altersgruppen vor: 2-4 Jahre (Kleinkinder), 5-7 Jahre (junge Kinder), 8-12 Jahre (Kinder) und 13-18 Jahre (Jugendliche).

Das Modul für Kleinkinder basiert auf den Angaben der Eltern, die Kinder selbst werden aufgrund ihres Entwicklungsstands (<5 Jahre) nicht befragt. Das PedQL-Modul (Eltern-Fremdbericht) ist geeignet, um zwischen gesunden Kindern und Kindern mit chronischen

Erkrankungen zu unterscheiden, und es wurde eine MID berichtet (4,5 Punkte für das generische Kernmodul, Fremdbericht durch die Eltern) [49]. Der Endpunkt wurde bei CLN2 nicht validiert.

Das PedsQL™-Modul „Familiäre Belastung“ („Family Impact“) umfasst 36 Fragen und 8 Dimensionen: Körperliche, emotionale, soziale und kognitive Funktion, Kommunikation, Sorgen, Alltagsaktivitäten und familiäre Beziehungen [50]. Im Rahmen der Studie 190-201/202 wurden fünf der 8 Dimensionen erhoben: Körperliche, emotionale, soziale und kognitive Funktion und Kommunikation [51]. Diese Dimensionen erfassen die jeweilige Situation der Eltern. Auf die Dimensionen Sorgen (aus Sicht der Eltern) und familiäre Situation (Alltagsaktivitäten und familiäre Beziehungen) wurde in der Studie 190-201 verzichtet [43].

Die in der Studie 190-201 eingesetzten PedsQL™-Module (Elternbericht für Kleinkinder und familiäre Belastung) entsprachen den zum Zeitpunkt der Planung der Studie gültigen Versionen (zur Operationalisierung siehe Tabelle 4-53). Hinsichtlich der Erhebung dieses Endpunkts gab es zwischen der Studie 190-201 und deren Folgestudie 190-202 keine Unterschiede. Der Endpunkt wurde in der Studie 190-201 zum natürlichen Verlauf nicht erhoben.

#### *CLN2-spezifischer Fragebogen*

Als Ergänzung zum PedsQL™ hat Biomarin einen CLN2-spezifischen Fragebogen neu entwickelt, der den Gesundheitszustand der Patienten krankheitsspezifisch erfasst. Das Instrument basiert auf Rückmeldungen aus Familien mit Kindern, die an CLN2 erkrankt waren. Hierzu wurden zwei Gruppen, davon eine in Europa und eine in den USA, befragt, welche Konsequenzen der Erkrankung sehr häufig auftreten und besonders belastend sind. Die Validität des Endpunkts ist nicht bekannt. Im Jahr 2015 fand eine Online-Umfrage statt, an der 24 CLN2-Experten (medizinische Fachkräfte und Patientenvertreter) teilnahmen, um Informationen zum derzeitigen Krankheitsmanagement zu erhalten. Die Experten stimmten darin überein, dass eine Verminderung folgenschwerer und potenziell lebensbedrohlicher Krampfanfälle ein wichtiges Ziel ist, um das Wohlbefinden und die QoL der Kinder so weit wie möglich zu erhalten [40].

Das Instrument umfasst 28 Items, die sechs Dimensionen abdecken: Krampfanfälle, Ernährung ohne/mit Gastrostomie-Schlauch, Schlaf, Verhalten und Alltagsaktivitäten, die jeweils auf einer 5-Punkt-Likertskala bewertet werden (0 = nie bis 4 = fast immer; siehe Tabelle 4-55). Die Eltern oder Erziehungsberechtigten füllen die Fragebögen aus. Die Erhebung des Endpunkts unterscheidet sich nicht zwischen der Studie 190-201 und der Folgestudie 190-202. Der Endpunkt wird in der Studie 190-201 zum natürlichen Verlauf nicht erhoben.

#### *EQ-5D (nicht eingeschlossen)*

Der EQ-5D wird nur in der Studie 190-202 erhoben. Da diese Studie noch nicht abgeschlossen ist und keine entsprechenden Daten aus der Studie 190-201 vorliegen, um den Verlauf über die Zeit zu charakterisieren, wird der Endpunkt bei der Nutzenbewertung nicht berücksichtigt.

*Schlussfolgerung:*

Die zur Messung der Lebensqualität eingesetzten Messinstrumente sind bis jetzt bei Patienten mit CLN2 oder deren Eltern/Betreuern nicht vollständig validiert und es liegen wenige Daten zum natürlichen Verlauf vor. Daher werden in dem vorliegenden Dossier deskriptive Längsschnittdaten aus der Studie 190-201 dargestellt.

Von wenigen Ausnahmen abgesehen sind nur Angaben durch die Patienten selbst ausreichend, um die Lebensqualität zu bestimmen, da nur die Patienten ihren eigenen Zustand subjektiv einschätzen können [52]. Die Erhebung der Lebensqualität bei Patienten mit CLN2 stellt eine besondere Herausforderung dar, insbesondere in Anbetracht des Alters der Patienten zu Studienbeginn ( $\geq 3$  Jahre). Daher konnte die Lebensqualität in der Studie 190-201/202 nur aus den Angaben der Eltern/Betreuer ermittelt werden; dies stimmt mit dem ISPOR-Report überein, der bei sehr jungen Kindern ( $< 5$  Jahre) eine Analyse auf Basis der Angaben der Eltern oder anderer Erwachsener empfiehlt. Diese sollte möglichst auf Beobachtungen beruhen und nicht auf die subjektiven Empfindungen oder Erfahrungen des Kindes ausgerichtet sein [47].

Eine Umfrage unter 19 Familien, die sich um Kinder mit CLN2 kümmern, zeigt, dass die Erkrankung in vielen Lebensbereichen mit hohen Belastungen verbunden ist (emotional-psychologisches Wohlbefinden, Beruf, Schulbildung, Finanzen und Gesundheit) [53]. Bei einer sehr schweren und rasch voranschreitenden Erkrankung wie der CLN2 mit einem hohen und mit der Zeit zunehmenden, speziellen Pflegebedarf werden auch die Auswirkungen auf die Eltern und Pflegepersonen als relevant erachtet. Entsprechend der IQWiG Methoden 4.2 können Interventionen auch Auswirkungen auf indirekt Betroffene wie zum Beispiel Familienangehörige und Pflegepersonen haben und diese können gegebenenfalls berücksichtigt werden [23]. Der CLN2-spezifische Fragebogen wurde entwickelt, um die spezifischen Probleme aus Sicht der Eltern zu erfassen. Die durch die Eltern berichteten QoL-Messinstrumente sind für die betroffenen Familien relevant, ihre Bedeutung für die Patienten ist allerdings unbekannt (unter Berücksichtigung der bekannten Herausforderungen bei sehr jungen, bereits behinderten Kindern).

***Verträglichkeit und Sicherheit****Patientenrelevanz*

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der *Verringerung von Nebenwirkungen* oder einer Verbesserung der Lebensqualität“ [32].

Nach einem Addendum des CONSORT-Statements (“Better Reporting of Harms in Randomized Trials”) können bei Populationen und Interventionen, die „harte“ Endpunkte verbessern (wie z. B. das Überleben), schwere und lebensbedrohliche UEs die einzigen sein, die für die Nutzen-Risiko-Bewertung von Bedeutung sind. Ein Therapieabbruch und Studienausschluss aufgrund eines UEs sind besonders wichtig, da sie die endgültige Entscheidung des Studienteilnehmers und/oder Arztes widerspiegeln, die Behandlung abzubrechen [54].

Gemäß IQWiG (Allgemeine Methodn 4.2) können dabei insbesondere diejenigen unerwünschten Wirkungen als relevant bezeichnet werden, die

- „den Nutzen der Intervention teilweise oder ganz aufwiegen könnten,
- sich zwischen zwei oder mehr ansonsten gleichwertigen Therapieoptionen erheblich unterscheiden könnten,
- insbesondere bei denjenigen Therapieoptionen auftreten, die besonders effektiv sein könnten,
- eine Dosis-Wirkungs-Beziehung haben könnten,
- von Patientinnen und Patienten als besonders wichtig angesehen werden und die
- von schwerwiegender Morbidität oder gar erhöhter Mortalität begleitet sind oder in der Folge mit einer erheblichen Beeinträchtigung der Lebensqualität einhergehen“ [23].

Auf Grundlage der Empfehlungen des IQWiGs werden diejenigen UEs als patientenrelevant erachtet, die den Nutzen der Intervention teilweise oder vollständig aufheben könnten und mit schwerwiegender Morbidität verbunden sind [23]. Bei einer sehr schweren und letztlich fatalen Erkrankung, für die keine krankheitsmodifizierende Therapie zur Verfügung steht, können nur schwere (CTCAE-Grad  $\geq 3$ ), schwerwiegende oder lebensbedrohliche UEs, UEs, die zum Therapieabbruch oder dauerhaften Beeinträchtigung führen, das Risiko-Nutzen-Verhältnis einer Intervention mit nachgewiesenem Nutzen beeinflussen. Dies entspricht dem Addendum des CONSORT-Statements [54].

Für die Eltern von Kindern mit CLN2 stellen Krampfanfälle ein großes Problem dar, und diese sind daher Bestandteil des durch Biomarin entwickelten CLN2-QoL-Fragebogens. Obwohl UEs, die den ICV-Port betreffen, nicht in direktem Zusammenhang mit Cerliponase alfa stehen, werden diese Komplikationen als patientenrelevant eingestuft, da eine Fehlfunktion oder Versagen des Ports einen chirurgischen Austausch erforderlich machen kann und z. B. Infektionen des Ports für diese besonders vulnerable und schwer kranke Zielpopulation ein hohes Risiko darstellen können. Es ist allerdings unklar, ob alle den ICV-Port betreffenden UEs (z. B. Problem mit dem Geräteanschluss) patientenrelevant sind.

Die Gesamtrate der UEs wird nicht eingeschlossen, da in der Operationalisierung der Nebenwirkungen auch Ereignisse wie z. B. Laborbefunde ohne klinische Symptome abgebildet sind, die nicht patientenrelevant sind. Die Gesamtrate der UEs wird lediglich ergänzend dargestellt

### *Validierung*

Die UEs wurden nach internationalen Standards in Übereinstimmung mit der GCP-Verordnung („Good Clinical Practice“) und relevanten, regulatorischen Leitlinien erhoben und berichtet.

### *Operationalisierung*

UEs von speziellem Interesse waren prospektiv im Studienprotokoll definiert (siehe Tabelle 4-57). Die Auswahl beruhte auf präklinischen Befunden, den bekannten Effekten von ERTs sowie einem Review der Literatur zu UEs, die in Zusammenhang mit ICV-Systemen berichtet wurden [37,55]. Die folgenden Endpunkte werden zur Sicherheit und Verträglichkeit dargestellt:

- Gesamtzahl der UEs
- Gesamtzahl der UEs pro Systemorganklasse (SOK)
- Sehr häufige UEs
- Alle schwerwiegenden und fatalen UEs
- Alle UEs CTCAE Grad  $\geq 3$
- Studienabbruch aufgrund eines UEs
- UEs von besonderem Interesse:
  - Überempfindlichkeits- und anaphylaktische Reaktionen
  - UEs, die mit der Infusion in Zusammenhang stehen („Infusion-associated reactions“, IAR oder „treatment-related events“, TRE)
  - UEs innerhalb der SOC „Erkrankungen des Nervensystems“, UEs, die durch die SMQ („Standardised MedDRA query“) „Konvulsionen“ identifiziert werden, sowie das ZNS betreffende UEs besonderen Interesses: Status epilepticus, Hydrozephalus, Meningitis
  - Unerwarteter, nicht anders erklärbarer und rascher Abfall des CLN2-Krankheits-scores
  - Den ICV-Port betreffende UEs

Der Schwere der UEs wurde nach CTCAE beurteilt und die UEs nach der zum Zeitpunkt der Analyse gültigen Version von MedDRA (18.1 oder 19.1) kodiert. Die Studie wurde in Übereinstimmung mit der GCP-Verordnung sowie relevanter Leitlinien durchgeführt und UEs entsprechend dieser Empfehlungen berichtet [26]. Die UEs von besonderem Interesse wurden durch MedDRA SMQs identifiziert oder entsprachen den MedDRA PTs („preferred terms“). Die Operationalisierung des Endpunkts ist in Tabelle 4-57 dargestellt.

Surrogatendpunkte wie Laborbefunde oder EEG (Elektroenzephalogramm)-Parameter wurden nicht berücksichtigt.

### Natürlicher Verlauf der Erkrankung

Das IQWiG („Allgemeine Methoden“) nimmt zur Situation bei extrem seltenen Erkrankungen wie folgt Stellung: „Für den Fall extrem seltener Erkrankungen oder extrem spezifischer Krankheitskonstellationen kann die Forderung nach (parallel) vergleichenden Studien unangemessen sein. Nichtsdestotrotz sind auch hier zumindest adäquate Dokumentationen des Verlaufs solcher Patientinnen und Patienten einschließlich des erwarteten Verlaufs ohne die ggf. zu prüfende Intervention (z. B. anhand historischer Patientenverläufe) möglich und bewertbar“ [23]. Um diesen Anforderungen gerecht zu werden, hat Biomarin zwei (teilweise überlappende) Analysen von Patientenregistern durchgeführt. Diese dienen der Charakterisierung des Krankheitsverlaufs sowie als historische Kontrolle für die Interventionsstudie mit Cerliponase alfa. Obwohl die Datensätze teilweise überlappen, werden die Ergebnisse beider Studien dargestellt, da die analysierten Endpunkte nicht identisch sind. Die aktualisierte Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901) [56] dient als historische Kontrolle der Studien 190-201/202. Bei dieser Analyse wurden spezifische Auswahlkriterien eingesetzt, um eine Patientenpopulation zu erhalten, die den wesentlichen Ein- und Ausschlusskriterien der Studie 190-201 entsprach.

Es wurden nur wenige quantitative Daten zum natürlichen Verlauf der CLN2 publiziert [5,6]. Eine Studie bei Patienten mit Wohnsitz in Deutschland oder der Schweiz (n=22) hatte zum Ziel, den Krankheitsverlauf mittels eines krankheitsspezifischen Bewertungsinstruments (Hamburg-Skala) quantitativ zu beschreiben und Informationen über den neurodegenerativen Krankheitsverlauf zu liefern (Steinfeld et al., 2002). Die Scores wurden bei denjenigen Patienten (n=16) bestimmt, die ausschließlich Träger der beiden häufigen Mutationen R208X und IVS 5-1G>C waren. Die meisten Patienten zeigten den typischen spätinfantilen Krankheitsverlauf, der durch einen sehr ähnlich verlaufenden Abbau motorischer und sprachlicher Fähigkeiten charakterisiert ist. Beide Fähigkeiten nahmen ab einem Alter von etwa drei Jahren rapide ab bis zum vollständigen Funktionsverlust im Alter von circa fünf Jahren. Bei zwei weiteren Patienten lagen sowohl der motorische als auch der Gesamtscore („total disability score“) außerhalb der 95igsten Perzentile und bei einem Patient wichen die motorischen und sprachlichen Scores sowie die Krampfneigung stark vom „Standardpatienten“ ab. Der Phänotyp unterschied sich zwischen den Patienten etwas, dennoch war der Krankheitsverlauf im Allgemeinen hoch vorhersehbar und entsprach dem klassisch spätinfantilen Krankheitsverlauf [5].

Bei 12 spanischen Kindern mit CLN2 war der Krankheitsverlauf sehr ähnlich (retrospektive Studie). Bei Auftreten der ersten Symptome waren die Kinder zwischen 1,6 und 3,7 Jahre alt und die Symptome umfassten eine verzögerte Sprachentwicklung und Fieberkrämpfe gefolgt von Epilepsie. Der Krankheitsverlauf wurde anhand der Krankenakten ermittelt und durch Angaben der Ärzte und Patienten ergänzt; eine spezifische Bewertungsskala wurde dabei nicht eingesetzt. Die Phase der Progression begann mit dem Verlust von Sätzen (im Median im Alter von 3,7 Jahren, 95% KI 3,1-4,3; 100% der Patienten) gefolgt von Gehunfähigkeit im Alter von etwa 4,5 Jahren (Median; 95% KI 4,3-4,6; 92% der Patienten) und Sprachunfähigkeit ebenfalls im Alter von 4,5 Jahren (Median; 95% KI 4,1-4,8; 92% der Patienten). Im Alter von fünf Jahren (Median) waren die Patienten blind und nicht mehr in der Lage, zu sitzen [4].

Tabelle 4-30 gibt einen Überblick der publizierten Literatur zum Alter bei Auftreten des ersten Krampfanfalls sowie erster Sprachstörungen, motorischer Einschränkungen und visueller Symptome und nachfolgender „Meilensteine“ der Erkrankung, die mit diesen Symptomen in Zusammenhang stehen.

Tabelle 4-30: Leitsymptome der Ceroid-Lipofuszinose Typ 2 -zeitlicher Verlauf

	<b>Nickel 2016 [1]</b>	<b>Pérez-Poyato 2013 [4]</b>	<b>Moore 2008 [7]</b>	<b>Worgall 2007 [6]</b>	<b>Steinfeld 2002 [5]</b>	<b>Chang, 2011 [3]</b>
Nationalität	Deutschland, Polen, Italien	Spanien	Kanada (Neufundland)	US, Kanada, England, Österreich, Spanien	Deutschland, Schweiz	Europa
n	77	12	28	18	16	17
Symptome im historischen Verlauf (Alter [Jahren])						
	Mittel (SD)	Median (95% KI)	Median (Max, Min)	Median (Max, Min)	Median	Median Max, Min
Erster epileptischer Anfall	3,4 (1,0)	3,4 (2,9-3,8) [partiell, n=10]	3,0 (1,8-4,0)	3,1 (2,0-6,3) [partiell, generalisiert]	2,7	3,0 (2,5-4,5) [partiell, generalisiert, n=16]
Sprachstörungen		3,7 (3,1-4,3) [Verlust der Fähigkeit zur Satzbildung]	-	3,0 (1,5-4,8) [inklusive verzögerter Sprachentwicklung]	3,1	2,3 (1-3,5) [inklusive verzögerter Sprachentwicklung, n=16]
Verlust motorischer Fähigkeiten		4,0 (3,9-4,1) [Ataxie]	-	3,5 (1,9-5,0)	3,1	3,5 (2,5-4,6) [Ataxie]
		4,5 (4,3-4,6) [Verlust der Gehfähigkeit]		5,7 (4,3-6,5) [Rollstuhlpflichtig]		5,1 (4,0-6,0) [Rollstuhlpflichtig, n=9]
Sehschwäche, Visusverlust		3,8 (3,2-4,4) [n=10]	4,0 (2,9-5,5) [n=22]	5,2 (4,0-7,7) [n=16]	4	5,0 (4,0-6,0) [n=8]
		5,1 (4,6-5,6) [Erblindung, n=11]				

	Nickel 2016 [1]	Pérez-Poyato 2013 [4]	Moore 2008 [7]	Worgall 2007 [6]	Steinfeld 2002 [5]	Chang, 2011 [3]
Diagnose	4,7 (1,6)	5,5 (2,0-7,0)		4,0 (2,0-5,3)		

Zusätzlich wurden die motorischen und sprachlichen Fähigkeiten in den Studien von Steinfeld (2002) und Nickel (2016) quantitativ erfasst.

Zusammen mit den Informationen aus neueren Reviews [2,3] ergibt sich ein sehr konsistenter Zeitverlauf der Erkrankung, der wie folgt beschrieben werden kann (siehe Abbildung 3):

- Krampfanfälle und Ataxien treten im Alter von 2-4 Jahren auf und diesen kann eine verzögerte Sprachentwicklung vorausgehen.
- *Die motorischen und sprachlichen Fähigkeiten nehmen ab einem Alter von etwa drei Jahren rasch und parallel zueinander ab und im Laufe der nächsten 2,5 Jahre gehen diese Fähigkeiten vollständig verloren.*
- Die Spastik der Gliedmaßen und der erniedrigte Muskeltonus der Rumpfmuskulatur führen im Alter von 4-6 Jahren dazu, dass die Patienten vollständig immobil werden.
- Die meisten Patienten können im Alter von sechs Jahren nicht mehr ohne Hilfestellung sitzen.
- Visusverluste können ab einem Alter von etwa vier Jahren auftreten, d. h. zu einem Zeitpunkt, zu dem die fortschreitenden, psychomotorischen Störungen bereits offensichtlich sind. Diese führen innerhalb von etwa drei Jahren zur Erblindung. Der Verlust des Sehvermögens unterscheidet sich zwischen den Patienten deutlicher und erfolgt langsamer als der Abbau motorischer und sprachlicher Fähigkeiten.
- Die meisten Kinder verlieren die Fähigkeit, zu schlucken und benötigen zur Ernährung unterstützende Maßnahmen durch eine Magensonde oder einen Gastrostomieschlauch.



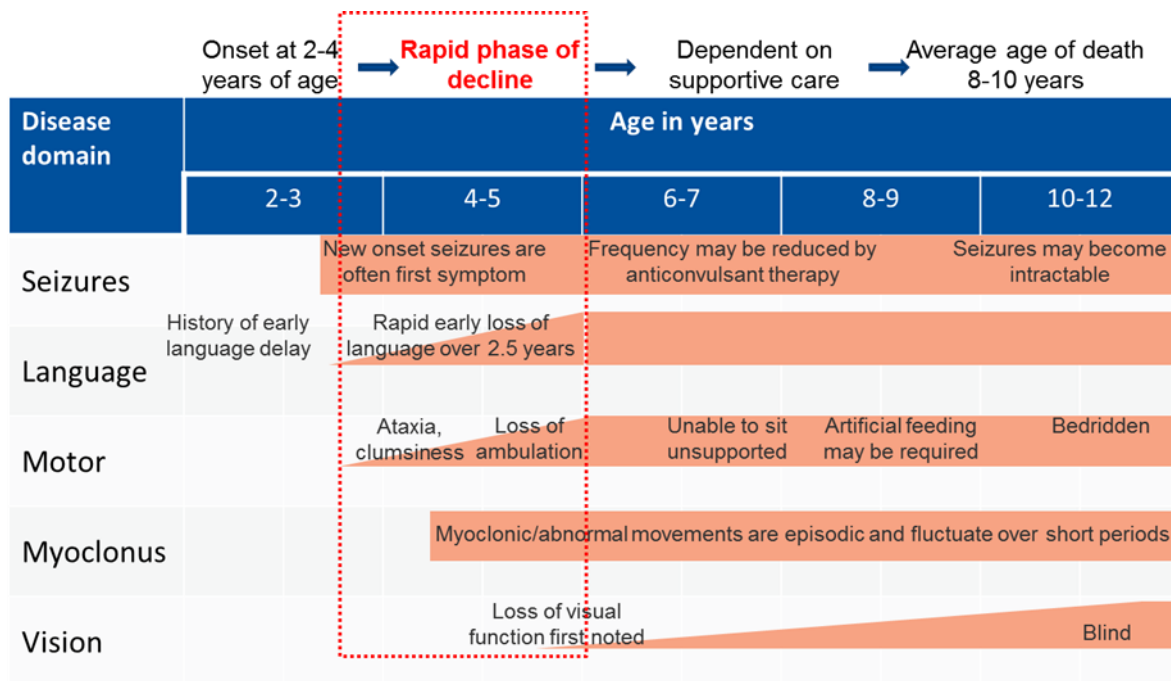


Abbildung 3: Typischer Verlauf der Ceroid-Lipofuszinose Typ 2

**Originalanalyse der Studie 190-901 zum natürlichen Verlauf [39]**

Bereits vor Beginn der Studie 190-201 hat BioMarin mit dem Universitätsklinikum in Hamburg-Eppendorf zusammengearbeitet, um die empirischen Daten zum natürlichen Verlauf zu bestätigen und zu erweitern. In die Analyse wurden Patienten aus zwei großen Kohorten in Hamburg (n=30; 29 ausgewertet, Längsschnittdaten) bzw. New York (n=49, Querschnittdaten; WCMC) eingeschlossen. Diese Analysen erweitern die quantitativen Ergebnisse zum natürlichen Verlauf der Originalstudien von Steinfeld et al. (2002) und Worgall et al. (2007) [5,6]. Ein wichtiges Ziel der Studie war es, den klinischen Verlauf der Krankheitsprogression zu bestätigen, und insbesondere, inwieweit der Funktionsverlust (motorisch und sprachlich) über die Zeit bei CLN2 vorhersagbar ist, gemessen mittels der Hamburg CLN2-Bewertungsskala (HML-Skala). Es handelt sich um eine retrospektive Analyse auf Basis klinischer Datenbanken.

Die Kohorte aus Hamburg wurde auf 30 Patienten mit genotypisch nachgewiesener CLN2 erweitert, deren Krankheitsverlauf unter Einsatz der krankheitsspezifischen, klinischen Bewertungsskala von Steinfeld et al. (2002) quantitativ erfasst wurde [5]. Diese Ergebnisse zum Krankheitsverlauf werden durch Querschnittdaten der WCMC-Kohorte (n=49) ergänzt. Bei der WCMC-Kohorte handelt es sich um die Screening-Daten aus zwei klinischen Studien (Kohorte 1 und 2). Mithilfe der Daten aus Hamburg sollte geklärt werden, wann die Erkrankung klinisch in Erscheinung tritt (Vorhersage mittels statistischer Methoden), wie schnell die Erkrankung voranschreitet und in welchem Alter Symptome auftreten bzw. die Diagnose gestellt wird. Die WCMC-Kohorte liefert quantitative Querschnittdaten bei der Erkrankung.

Die Studie 190-901 hatte folgende Ziele:

- Bestimmung der Progressionsrate und deren Variabilität unter Einsatz einer krankheits-spezifischen Skala nach Auftreten klinischer Symptome
- Bestimmung des Zeitpunkts und der Variabilität der ersten klinischen Symptome (Krankheitsbeginn)
- Beschreibung der ersten klinischen Symptome
- Beschreibung der Altersverteilung zum Zeitpunkt der Diagnose
- Beschreibung des Zusammenhangs zwischen Genotyp und klinischem Verlauf
- Modellierung des Zusammenhangs zwischen den MRI-Befunden und der Schwere der Erkrankung sowie dem Alter der Patienten
- Vergleich des Patientenalters und der Krankheitsschwere bei zwei unabhängigen Patientenkohorten

Der klinische Verlauf der Erkrankung wurde mittels der Hamburg- und der WCMC CLN2-Bewertungsskalen quantifiziert, die jeweils aus vier Domänen bestehen (siehe Tabelle 4-28). Die ML-Subskala (0-6 Punkte), die der Hamburg- und WCMC-Skala gemeinsam ist, stellte den primären Endpunkt dar. Analysiert wurden die intraindividuellen Längsschnittdaten zum klinischen Verlauf (unter Einsatz der HML-Skala) der Patienten aus Hamburg. Zusätzlich wurde untersucht, inwieweit die Daten des WCMC mit denjenigen aus Hamburg übereinstimmen; hierzu wurden die WCMC-Scores zur motorischen und Sprachfunktion (klinische Schwere getrennt nach Alter) mit den Längsschnittdaten aus Hamburg (HML) verglichen. Bei 31 der insgesamt 49 Patienten der WCMC-Kohorte wurde die WCMC-Skala (Gang/Sprache) und bei den verbleibenden 18/49 Patienten die HML-Skala eingesetzt. Zusätzlich wurden der Gesamtscore (0-12 Punkte) und beide Subskalen getrennt ausgewertet.

Bei der Kohorte aus Hamburg handelt es sich um die bislang umfangreichste Längsuntersuchung des klinischen Phänotyps unter Einsatz einer klinischen Bewertungsskala. Zwischen dem ersten Auftreten von Symptomen und der Diagnose CLN2 vergeht oft eine beträchtliche Zeit. Bei den Daten, die sich auf den Zeitraum vor Diagnose der CLN2 beziehen, handelt es sich um retrospektive Angaben, während die meisten Werte nach Diagnosestellung bei den Patienten aus Hamburg tatsächlich gemessen wurden; bei der Analyse wurde nicht zwischen diesen Datensätzen unterschieden. Die retrospektiven Daten werden als valide eingeschätzt, da jede Punktänderung auf der HML-Skala einen klinischen „Meilenstein“ darstellt, an den sich die Eltern oder Betreuer gut erinnern dürften.

Die WCMC-Daten bestanden in der Regel aus nur einem Messwert oder gelegentlich zwei Werten, die kurz hintereinander erhoben worden waren. Bei Patienten mit zwei Werten wurde die erste Beobachtung analysiert. Bei den Patienten der WCMC-Kohorte 1 (n=31/49) wurde der Mittelwert aus drei unabhängigen Bewertungen gebildet.

Um die Rate der Krankheitsprogression abzuschätzen, wurden in der Primäranalyse der HML-Skala (0-6 Punkte) zwei Methoden eingesetzt: (1) Neigung („Slope“) der Linie zwischen dem Beginn und Ende der progressiven („decline“) Phase und (2) Neigung („Slope“), ermittelt durch lineare Regression über alle Messungen während der Phase der Progression („decline“):

- Methode (1): Alle Patienten der Kohorte aus Hamburg hatten initial einen ML-Summenscore von sechs und mindestens einen Wert, bei dem die Summe Null betrug. Die Rate der Progression wurde bei jedem Patienten geschätzt, indem der Wert von sechs durch die Zahl der Monate seit der letzten Messung mit einem Wert von sechs bis zur ersten Erhebung mit einer Punktzahl von Null geteilt wurde. Zur Schätzung der Zahl der Monate wurde der mittlere Wert zwischen der letzten Beobachtung mit einer Punktzahl von sechs und der anschließenden Beobachtung als Beginn der „Decline“-Phase eingestuft. Gleichmaßen war das Ende der „Decline“-Phase als mittlerer Wert zwischen dem ersten Wert von Null und der vorherigen Beobachtung definiert. Die daraus resultierende Neigung („Slope“) wurde mit 12 multipliziert und als mittlere Änderung über 12 Monate ausgedrückt.
- Methode (2): Das Fortschreiten der Erkrankung wurde bei jedem Patienten mittels linearer Regression analysiert (d. h. über den Zeitraum zwischen der letzten Beobachtung mit einem Wert von sechs und dem ersten Punktwert von Null).

Der Datensatz aus Hamburg enthielt ein eineiiges Zwillingsspaar mit identischen Daten und nur einer der Zwillinge wurde in die Analyse aufgenommen, was zu einer Stichprobengröße von 29 Patienten führte.

Die WCMC-Kohorte bestand aus 49 Patienten mit bestätigter CLN2 und umfasst klinische sowie MRT-Daten. Nahezu alle Patienten waren kaukasischer Herkunft (n=44/49, 90%; Daten nur für die WCMC-Kohorte vorhanden), die Hälfte der Kohorte war männlich (n=38/79, 48%). Die Mutationen c.622C>T und c.509-1G>C waren die häufigsten Krankheitsallele (66% bzw. 45% Hamburg-Kohorte). Bei nur 14% der Patienten aus Hamburg wurde keines der häufigen Krankheitsallele nachgewiesen und bei 31% andere Mutationen. Erste klinische Symptome traten im Alter von 12-53 Monaten (n=21 mit Daten aus Hamburg) und im Mittel 34,8 Monaten (entsprechend etwa drei Jahre) auf. Der Interquartilsabstand (25igste bis 75igste Perzentile) schwankte zwischen 30 (2,5 Jahre) und 40 Monaten (etwa 3,5 Jahre). In Anbetracht des mittleren Alters bei Auftreten erster Symptome (35 Monate) und der Diagnosestellung (63 Monate), vergingen zwischen den ersten Symptomen und der Diagnose CLN2 fast 2,5 Jahre (28 Monate). Bei 22 Patienten aus Hamburg mit entsprechenden Daten stellten Krampfanfälle (73%), Sprachprobleme (59%) und motorische Störungen (45%) die häufigsten klinischen Erstmanifestationen dar.

Die Progressionsrate, gemessen auf der HML-Skala (0-6 Punkte), wurde bei den Patienten aus Hamburg (n=29) unter Einsatz von zwei verschiedenen Methoden geschätzt:

- Die Slope-Analyse (berechnet aus der Linie zwischen zwei Punkten) ergab eine mittlere (SD) Progressionsrate von -2,89 (1,36) Punkten/Jahr (Bereich -6,3, -1,0).

- Berechnet als lineare Regression betrug die mittlere (SD) Progressionsrate -2,18 (1,07) Punkte/Jahr (Bereich -4,8, -0,2).

In Abbildung 4 sind Querschnittsdaten zur Progressionsrate bei den Patienten aus Hamburg getrennt nach Alter (Geburt bis 12 Jahre) dargestellt. Die Auswertungen der klinischen Scores zeigen übereinstimmend, dass die Erkrankung innerhalb eines Zeitraums von etwa 2,5 Jahren rasch von weitgehend altersgerecht normal bis hin zum nahezu vollständigen Ausfall motorischer Fähigkeiten und der Sprache voranschreitet. Betrachtet man die individuellen Scores der Patienten aus Hamburg, so ist der Krankheitsverlauf bei den meisten Patienten hoch vorhersehbar, jeweils mit ersten Symptomen im Alter zwischen zwei und vier Jahren gefolgt von einem Progress von altersgerecht normal bis zum vollständigen Verlust motorischer und sprachlicher Fähigkeiten zwischen fünf und sieben Jahren. Außenseiter waren zu Beginn der Erkrankung in der Regel älter und die Erkrankung schritt in frühen Stadien weniger stark voran; dennoch zeigten auch diese Patienten im fortgeschrittenen Stadium den typischen, durch rasch zunehmenden Funktionsverlust gekennzeichneten Verlauf

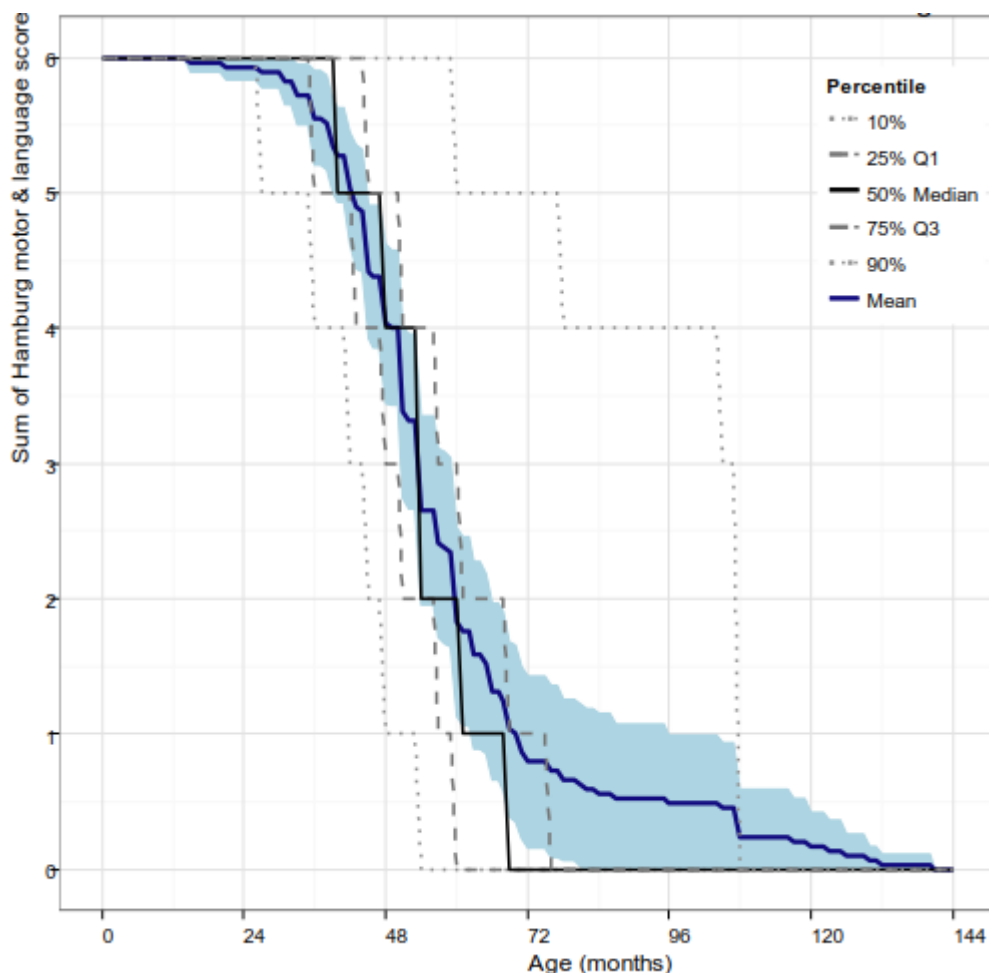


Abbildung 4: HML-Score nach Alter bei Patienten der Hamburg Kohorte (n=29)

Die schattierten Bereiche entsprechen den 95% KIs.

Natural History Study 190-901: Figure 9.1.1 [39]

Die mittlere (SD) Progressionsrate bezogen auf den Gesamtscore der Hamburg-Skala (0-12 Punkte) betrug bei den Patienten aus Hamburg -2,26 (1,47) Punkte pro Jahre und schwankte zwischen -8,4 und -0,5 Punkten pro Jahr.

In Abbildung 5 ist die Änderung der motorischen bzw. Sprach-Subskalen (Hamburg-Skala, jeweils 0-3 Punkte) getrennt nach Alter auf Basis von Querschnittsdaten dargestellt. Beide Skalen zeigten einen ähnlich progredienten Verlauf. Der Visusverlust (Abbildung 6) trat innerhalb des Krankheitskontinuums etwas später auf als der Abbau motorischer und sprachlicher Fähigkeiten. Grand mal-Anfälle, gemessen auf der Hamburg-Subskala, unterlagen häufig Fluktuationen und zeigten mit zunehmendem Alter einen U-förmigen Verlauf (Abbildung 7).

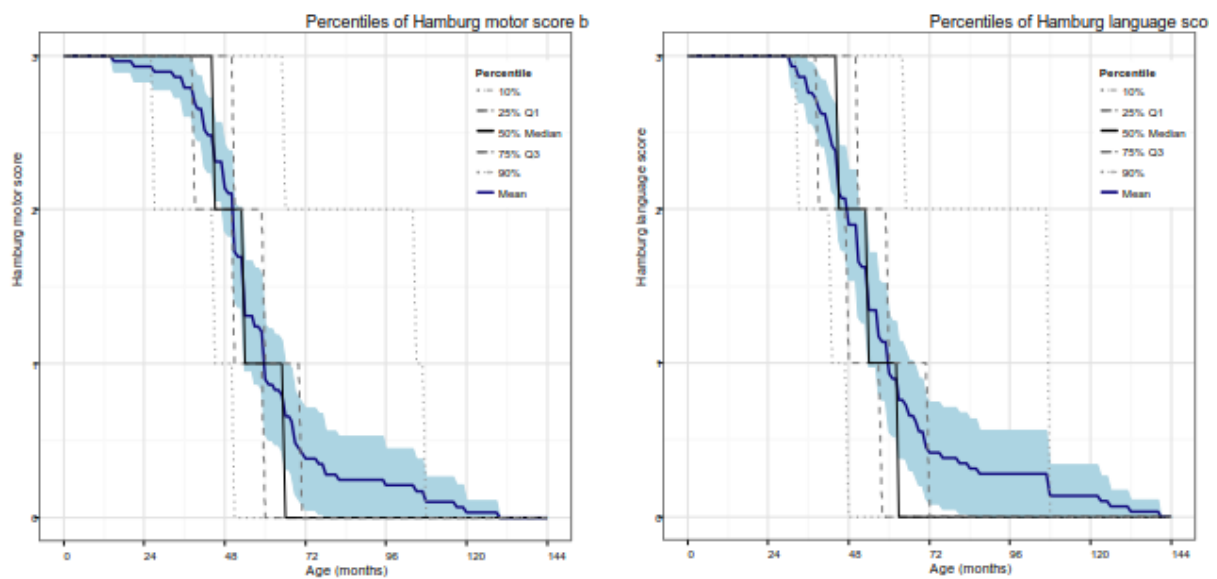


Abbildung 5: Querschnittsanalyse der motorischen (links) und Sprachdomäne (rechts) der HML-Skala, Hamburg Kohorte (n=29)  
Natural History Study 190-901: Figure 13.4.1.1 [39]

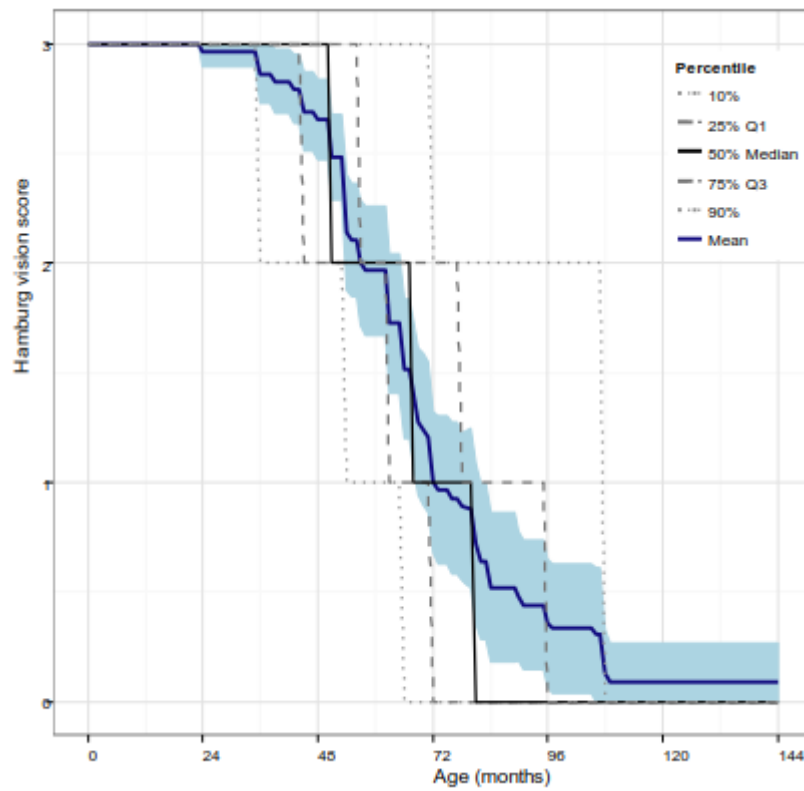


Abbildung 6: Querschnittsanalyse der Domäne „Sehvermögen“ der HML-Skala, Hamburg Kohorte (n=29)

Natural History Study 190-901: Figure 13.4.2.1 [39]

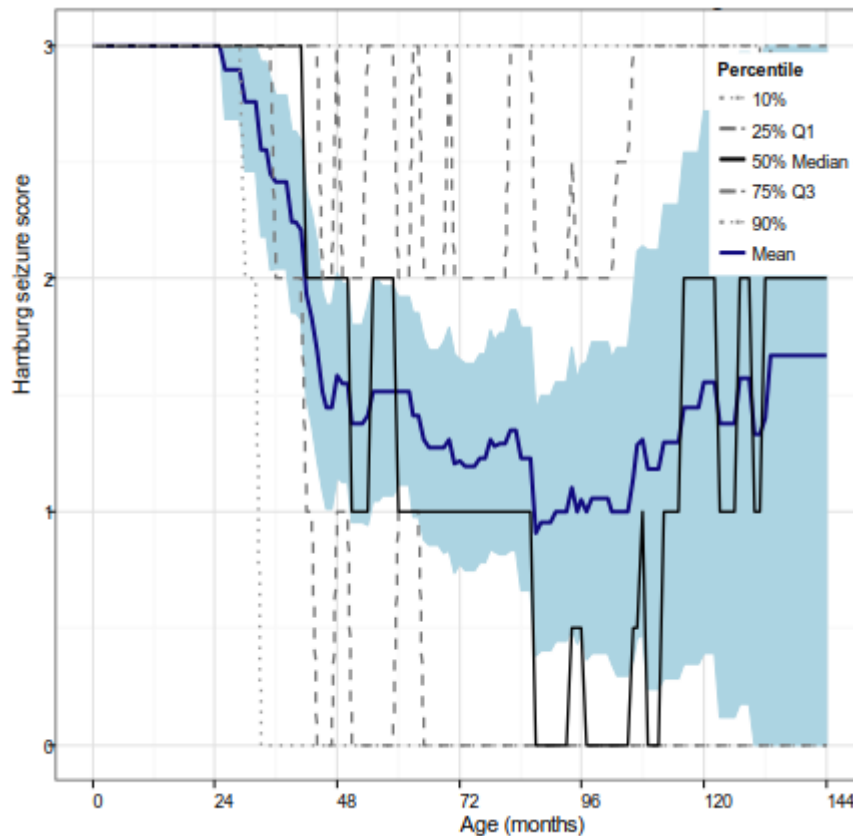


Abbildung 7: Querschnittsanalyse der Domäne “Krampfanfälle” der HML-Skala, Hamburg Kohorte (n=29)

Natural History Study 190-901: Figure 13.4.3.1 [39]

Die Longitudinaldaten aus Hamburg (HML-Summenscore, n=29) und der WCMC-Kohorten (Kohorte 1, WCMC Gang/Sprache-Summenscore, n=31; Kohorte 2, Hamburg-Score, n=18) wurden graphisch übereinander projiziert, um den Zusammenhang zwischen dem Alter der Patienten und dem Schweregrad bei zwei unabhängigen Patientengruppen zu untersuchen. Die meisten Werte des WCMC lagen innerhalb des Interquartil-Bereichs der Kohorte aus Hamburg (zum Vergleich siehe Abbildung 4). Diese Daten zeigen bei *zwei großen, unabhängigen Patientengruppen* und unter Einsatz von zwei sehr ähnlichen Erhebungsinstrumenten einen auffällig ähnlichen Zusammenhang zwischen der Schwere der Erkrankung und dem Alter der Patienten (Abbildung 8).

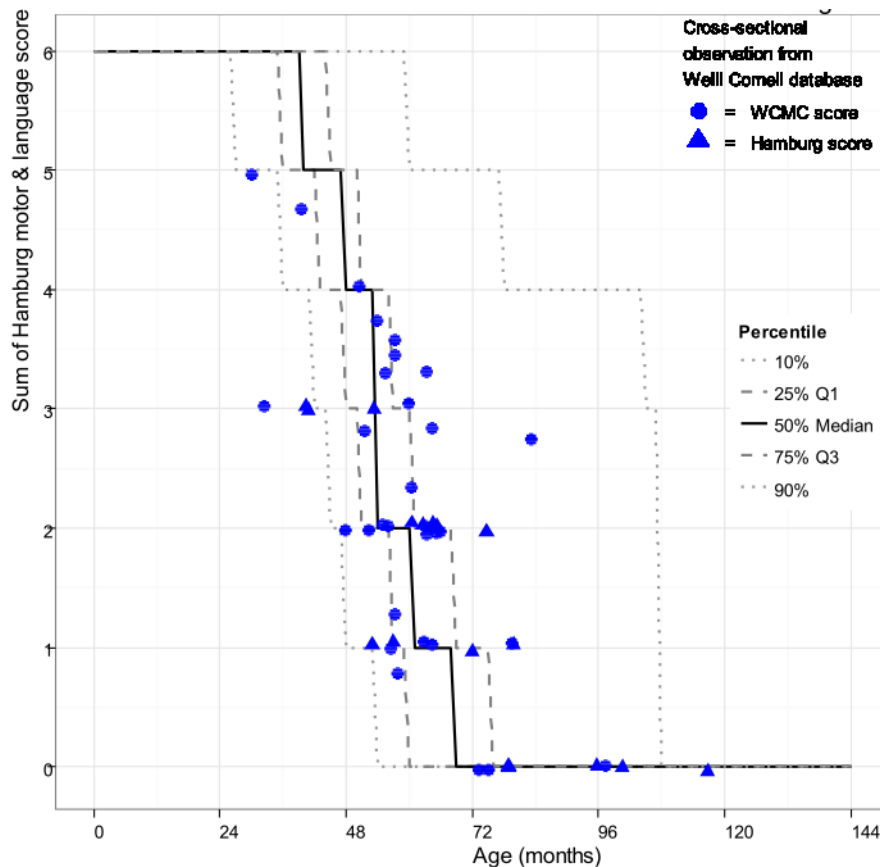


Abbildung 8: Summenscore der kombinierten motorischen Skala und Sprachskala in unabhängigen Patienten-Kohorten, gemessen mittels unterschiedlicher Instrumente: WCMC-Kohorten und Hamburg Kohorte

Bei Patienten der WCMC Kohorte 1 (n=31) wurde der WCMC-Summenscore durch drei Personen unabhängig voneinander bewertet (Kreise), während bei WCMC Kohorte 2 (n=18) der Hamburg-Score erhoben wurde (Dreiecke).

Für die Hamburg Population aus dem DEM-CHILD Register sind der Verlauf des Medians und die Perzentile dargestellt (siehe Abbildung 4)

Natural History Study 190-901: Figure 9.3.1 [39]

Nach Auftreten erster Symptome (meist Krampfanfälle in 73% der Fälle) ist der weitere Krankheitsverlauf durch rasch voranschreitende, neurologische Defizite gekennzeichnet und entspricht dem gut bekannten, klassisch spät-infantilen Phänotyp. Nach Ausbruch der Erkrankung verschlechterte sich der HML-Score jährlich um 2,2 Punkte. Obwohl es sehr wenige Ausreißer gibt, schreitet die Erkrankung bei den meisten Patienten innerhalb von drei Jahren nach Beginn der „Decline“-Phase in hoch vorhersehbarer Weise von weitgehend normalen Ausgangswerten (Wert von 6) bis zum nahezu vollständigen Funktionsverlust der sprachlichen und motorischen Fähigkeiten (Wert von 0) voran. Der Krankheitsverlauf der Patienten aus Hamburg (Längsschnittdaten) war demjenigen der unabhängigen WCMC-Kohorte (Querschnittsdaten) sehr ähnlich.



***Aktualisierte Analyse der Studie 190-901 zum natürlichen Verlauf der Erkrankung [56]***

Um weitergehende Erkenntnisse zum klinischen Verlauf von NCL-Erkrankungen zu erhalten, wurde ein europaweites Patientenregister mit mehreren Studienzentren eingerichtet (eine EU-FP7 gesponsertes Forschungsprogramm). Bei DEM-CHILD handelt es sich um eine fortlaufende, multizentrische und multinationale klinische Datenbank mit Sitz in Hamburg, die - verglichen mit dem Originalbericht der Studie 190-901 - aktualisierte klinische Daten enthält (Datenschnitt Februar 2015). Die aktualisierte Analyse umfasst Patienten aus zwei Studienzentren (Hamburg, Deutschland und Verona, Italien). Die Methodik wurde nach Rücksprache mit den Zulassungsbehörden verfeinert. Zudem hatte dies zum Ziel, die Analysen denjenigen der klinischen Studie 190-201 anzupassen. BioMarin hatte keinen Einfluss auf die Datenbankeinträge in den Studienzentren und alle CLN2-Patienten, die zum Zeitpunkt des Datentransfers (Februar 2015) an DEM-CHILD teilnahmen, wurden in der Analyse eingeschlossen.

Das Design der Studie ist in Tabelle 4-77 beschrieben, um derzeit gültigen Berichtsstandards zu genügen.

Die aktualisierte Analyse (Supplemental Report der Studie 190-901) hatte folgende Ziele:

- Charakterisierung des natürlichen Verlaufs bei Patienten des DEM-CHILD-Programms, um als historische Kontrolle der Krankheitsprogression für Interventionsstudien mit Cerliponase alfa zu dienen.
- Analyse unter Einsatz einer krankheitsspezifischen Bewertungsskala in dem größeren DEM-CHILD-Datensatz, wie schnell und wie variabel die Erkrankung fortschreitet nachdem die Patienten symptomatisch geworden sind
- Deskriptive Statistik der Patienten-Charakteristika.

Vor der Analyse der HML-Daten wurden verschiedene Selektionskriterien (Filter) eingesetzt, um eine Patientenpopulation zu erreichen, die für einen historischen Vergleich der Krankheitsprogression mit den Teilnehmern der Studie 190-201 geeignet ist: Mindestens zwei HML-Werte, mindestens zwei HML-Werte im Alter von  $\geq 36$  Monaten, mindestens ein HML-Wert von  $\geq 3$  Punkten, mindestens zwei HML-Werte von  $\geq 3$  Punkten, mindestens zwei HML-Werte zwischen einem und fünf Punkten und mindestens ein HML-Wert mindestens sechs Monate nach dem ersten HML-Wert (Tabelle 4-31).

Tabelle 4-31: Auswertbare Patienten des DEM-CHILD Registers

	<b>Hamburg (n = 63)</b>	<b>Verona (n = 11)</b>	<b>Overall (n = 74)</b>
Patienten nach Ausschluss eines Zwillings	62 (98%)	11 (100%)	73 (99%)
Patienten mit mindestens einem HML-Wert	47 (75%)	11 (100%)	58 (78%)
Patienten mit mindestens zwei HML-Werten	42 (67%)	9 (82%)	51 (69%)
Zwei HML-Werte im Alter von $\geq 36$ Monaten	42 (67%)	9 (82%)	51 (69%)
Ein HML-Wert von $\geq 3$ Punkten	40 (63%)	9 (82%)	49 (66%)
Zwei HML-Werte zwischen ein und fünf Punkten	40 (63%)	8 (73%)	48 (65%)
Mindestens ein HML-Wert $\geq 6$ Monate nach der ersten Erhebung der HML-Skala	33 (52%)	8 (73%)	41 (55%)

Von den insgesamt 74 Patienten mit CLN2 stammten 63 aus dem Zentrum in Hamburg und 11 aus dem Zentrum in Verona. Nach Einsatz der HML-Filter und Ausschluss eines Zwillings verblieben 41/74 Patienten (55%; auswertbare Patienten) in der Analyse und erfüllten die Eignungskriterien für den historischen Vergleich mit den Patienten der Studie 190-201. Von den 41 Patienten, die in die erweiterte Analyse der Studie 190-901 eingeschlossen wurden, handelte bei 26 um Patienten des ursprünglichen Reports zur Studie 190-901 und bei 15 um neu aufgenommene Patienten (7 in Hamburg und 8 in Verona). Die meisten Patienten waren zwischen 1990-1999 (n=19/41, 46%) und 2000-2009 (n=16/41, 39%) geboren worden und nur sechs Patienten zwischen 1960 und 1989 (15%). Die meisten der Patienten waren männlichen Geschlechts (59%; weiblich: 32%, unbekannt: 10%; Supplemental Report: Tabelle 8.1). Erste Anzeichen und Symptome, meist Krampfanfälle und Sprachprobleme, traten im Alter von etwa drei Jahren auf. Die Diagnose wurde etwa zum Zeitpunkt des fünften Geburtstags gestellt, d. h. fast zwei Jahre nachdem die Patienten erstmalig symptomatisch geworden waren. Bei sechs Patienten (15%) wurden zwei eher seltene Krankheitsallele nachgewiesen (siehe Tabelle 4-36).

Die Änderung, gemessen auf der HML-Skala, stellte den primären Endpunkt des erweiterten Reports der Studie 190-901 dar. Die Progressionsrate über 48 Wochen wurde anhand der Neigung („Slope“) der HML-Messwerte über den Zeitraum der „Decline“-Phase berechnet. Wie bereits in der ursprünglichen Analyse wurden zwei Algorithmen eingesetzt (erster/letzter Wert und lineare Regression; siehe oben).

Zur Überprüfung der Sensitivität der Slope-Analyse, wurde ein gemischtes Modell mit Messwiederholung (MMRM) eingesetzt: Hierbei wurden die Daten unter verschiedenen Varianzannahmen im Abstand von sechs Monaten ab einem Alter von 36 Monaten sowie dem Alter bei Diagnosestellung analysiert. Da die HML-Werte nicht exakt im Abstand von sechs Monaten erhoben wurden, war für die Werte nach Baseline ein Zeitfenster ( $\pm 1$  Monat) definiert. Falls zu Baseline (d. h. entweder Alter von 36 Monaten oder Alter bei Diagnose) keine Werte innerhalb eines Monats vorlagen, wurde die LOCB-Methode („last observation carried backwards“) genutzt, um Baseline festzulegen. Die Daten wurden bis zu sechs Monate nach

der letzten Beobachtung unter Einsatz von LOCF („last observation carried forward“) und bis zu sechs Monate vor der ersten Beobachtung mittels LOCB extrapoliert.

Zusätzlich wurde berechnet, wie lange bei den Patienten der HML-Score erhalten blieb (jeweils Verbleib auf benachbarten Punktwerten wie z. B. [5-4], [4-3], [3-2]), d. h. es wurde der Zeitraum zwischen dem ersten höheren Wert des Paares und dem Zeitpunkt, zu dem der Wert um zwei (oder mehr) Punkte niedriger war, kalkuliert. Bei der Analyse wurde davon ausgegangen, dass der Abfall nicht linear verläuft.

Da nur sehr wenige MRI-Ergebnisse in dem Datensatz vorhanden waren (6 Patienten), konnten diese nicht ausgewertet werden.

Der mittlere Abfall (SD) des HML-Scores betrug bei den 41 auswertbaren Patienten unter Einsatz des ersten Algorithmus (erster/letzter Wert) 2,09 (0,966) Punkte über 48 Wochen (95% KI 1,70, 2,40) und 2,09 (0,988) Punkte über 48 Wochen (95% KI 1,78, 2,40), berechnet mittels linearer Regression. Die Progressionsrate war bei den Patienten aus Verona etwas niedriger als in Hamburg (1,54 Punkte/48 Wochen verglichen mit 2,23 Punkten/48 Wochen; Tabelle 4-32). Diese Variabilität beruht möglicherweise darauf, dass nur wenige Patienten in Verona teilnahmen und bei diesen auch nur relativ wenige Beobachtungen vorlagen (bei 6/8 Patienten nur zwei Werte).

Tabelle 4-32: Abfall (Rate) des HML-Scores bei unbehandelten CLN2-Patienten (Punkte/48 Wochen; auswertbare Patienten) [56]

Analyse	Mittelwert $\pm$ SD [95% KI]		
	Hamburg (n=33)	Verona (n=8)	Overall (n=41)
Algorithmus: Erste /letzte Erhebung	2,23 (0,973) [1,88, 2,57]	1,54 (0,758) [0,91, 2,17]	2,09 (0,966) [1,79, 2,40]
Lineare Regression	2,22 (1,003) [1,87, 2,58]	1,55 (0,754) [0,92, 2,18]	2,09 (0,988) [1,78, 2,40]
Auswertbare Patienten: Patienten, die den Eignungskriterien der Studie 190-201 weitgehend entsprachen. 190-901 Supplemental Report: Tabelle 9.1, Tabelle 14.2.1, Tabelle 14.2.2			

Als Sensitivitätsanalyse wurde ein MMRM-Modell eingesetzt, in dem die HML-Werte im Abstand von sechs Monaten unter verschiedenen Varianzannahmen ab einem Alter von 36 Monaten oder bei Diagnose untersucht wurden. Die Progressionsrate (Abfall des HML-Scores) schwankte in diesem Modell zwischen -1,286 und -1,457 Punkten/48 Wochen (Tabelle 4-33) und ist somit niedriger als die oben angegebenen Werte. Die unterschiedlichen, zur Imputation fehlender Werte eingesetzten Methoden und Annahmen könnten zu diesem Unterschied bzw. der geringeren Progressionsrate beigetragen haben:

- Zur Imputation wurden in dem MMRM-Modell konservative Methoden eingesetzt: LOCF und LOCB;

- Die Daten wurden bis zum ersten ML-Wert von 0 analysiert;
- Bei den Analysen ab einem Alter von 36 Monaten ist zu beachten, dass viele Patienten zu diesem Zeitpunkt einen ML-Wert von sechs hatten.
- Bei den Analysen ab dem Zeitpunkt (Alter) bei Diagnosestellung nahm der Abfall des ML-Scores von 1 auf 0 Punkte einen relativ hohen Anteil innerhalb der Nachbeobachtung ein (dieser verläuft relativ betrachtet langsamer als der Abfall von fünf auf einen Punkt).

Tabelle 4-33: MMRM-Analysen des Abfalls (Rate) des HML-Scores bei unbehandelten CLN2-Patienten (Punkte/48 Wochen; auswertbare Patienten) [56]

Analyse	Rate/48 Wochen	95% KI	p-Wert <sup>1</sup>
MMRM: Zeit nach Diagnose (bis zu 72 Monate), heterogene, autoregressive Varianz <sup>2</sup>	1,286	(1,03; 1,54)	<0,0001
MMRM: Zeit nach Diagnose (0-30 Monate), unstrukturierte Varianz <sup>2</sup>	1,457	(1,12; 1,79)	<0,0001
MMRM: Nachbeobachtung ab einem Alter von 36-144 Monaten, heterogene, autoregressive Varianz <sup>3</sup>	1,286	(1,11; 1,44)	<0,0001
MMRM: Nachbeobachtung ab einem Alter von 36-72 Monaten, unstrukturierte Varianz <sup>3</sup>	1,395	(1,16; 1,63)	<0,0001

<sup>1</sup> Test nach Wald:  $H_0$  Rate = 0 vs.  $H_A$  Rate  $\neq$  0  
<sup>2</sup> ML-Score in 6-Monatsintervallen nach Diagnosestellung (6, 12, 18 und 72 Monaten bzw. 6, 12, 18, 24 und 30 Monate) unter Einsatz eines MMRM-Modells mit linearem Trend; adjustiert nach Alter bei Diagnosestellung und ML-Score zum Zeitpunkt der Diagnose.  
<sup>3</sup> ML-Score im Alter von 36, 42, 48 bis 144 Monate bzw. 36, 42, 48 bis 72 Monate unter Einsatz eines MMRM-Modells mit linearem Trend.  
190-901 Supplemental Report: Tabelle 9.2, Tabelle 14.2.7.1, Tabelle 14.2.7.2, Tabelle 14.2.7.3

Abbildung 9 zeigt die individuelle, lineare Abnahme des ML-Scores von fünf auf einen Punkt (lineare Phase der klinischen Progression). Die Abnahme von sechs auf fünf Punkte sowie einen auf null Punkte wurde hierbei nicht berücksichtigt, da diese nicht linear verlaufen. Die motorischen und sprachlichen Fähigkeiten nehmen bei den Patienten der historischen Kontrollgruppe nach Auftreten von Symptomen in vorhersagbarer Weise ab. Das Alter zu Beginn der Erkrankung schwankt zwischen den Patienten etwas. Dennoch zeigen die Slope-Analysen einen ausnahmslos gleichgerichteten Verlauf, d. h. die Fähigkeiten nehmen bei allen Patienten deutlich ab. Die Analyse der individuellen Verläufe legt nahe, dass die Erkrankung rasch voranschreitet und hoch vorhersagbar verläuft.

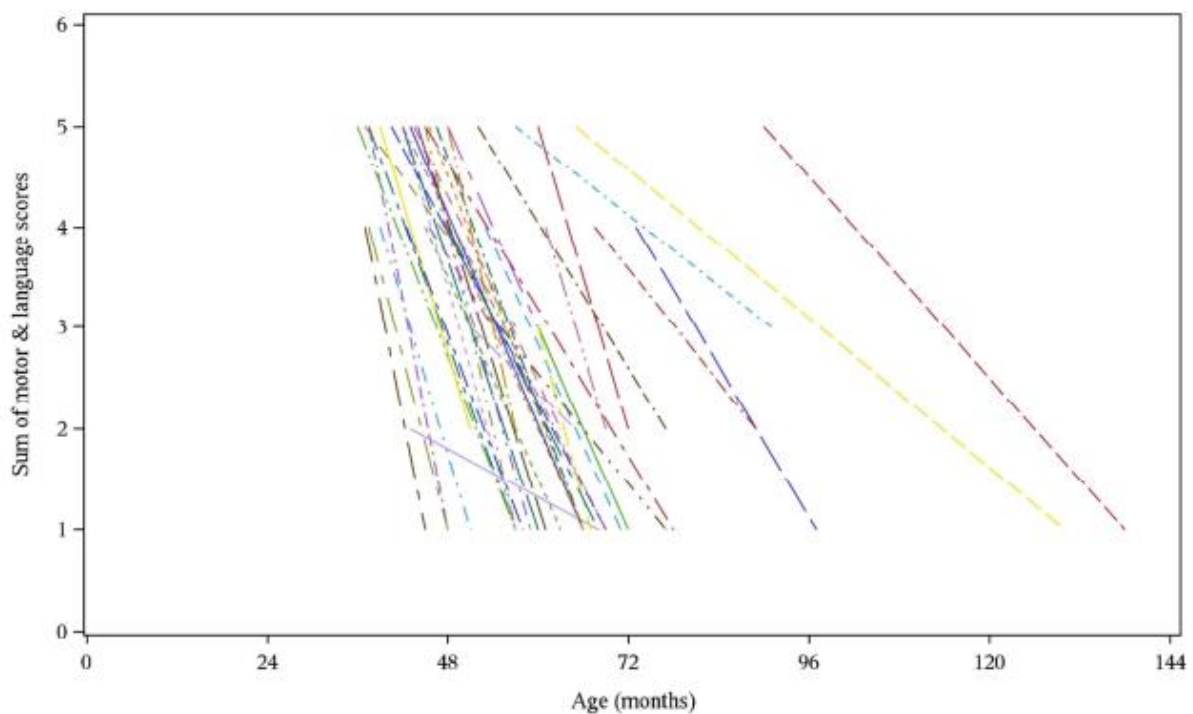


Abbildung 9: Individueller Abfall des HML-Scores (Punkte) während der linearen Phase der Krankheitsprogression (Schätzung unter Einsatz des Algorithmus erste/letzte Erhebung; auswertbare Patienten) [56]

Der lineare Abfall der ML-Scores wurde zwischen 5 und 1 Punkten geschätzt. 190-901 Supplemental Report: Abbildung 14.2.4

Zusätzlich wurde analysiert, wie lange ein bestimmter HML-Wertebereich bei den Patienten erhalten bleibt (Verweildauer auf einem Score von 5-4, 4-3, 3-2 und 2-1). Bei Patienten mit ML-Werten von 4-3, 3-2 und 2-1 Punkten betrug die mittlere Zeit bis sich der Score um zwei Punkte verschlechtert hatte jeweils weniger als ein Jahr (6,4 bis 11,4 Monate) und auch im Median weniger als ein Jahr (6,0 bis 9,0 Monate). Bei ML-Werten von 5-4 Punkten dauerte es etwas länger bis der Wert um zwei Punkte abgenommen hatte. Die Verweildauer betrug bei diesen Patienten im Mittel 13,9 Monate und im Median 14,0 Monate (190-901 Supplemental Report: Tabelle 9.3).

Die aktuellen Ergebnisse stützen den ursprünglichen Bericht der Studie 190-901: Die Erkrankung schreitet schnell voran und ihr Verlauf ist weitestgehend vorhersagbar. In die historische Kontrollgruppe wurden nur Patienten der Studie 190-901 eingeschlossen, die bestimmte Eignungskriterien erfüllten. Diese dienten dazu, eine Patientenpopulation zu identifizieren, die den Teilnehmern der Studie 190-201/202 ähnlich ist (Tabelle 4-31)

#### 4.4.2.2 Charakteristika der eingeschlossenen Studien

Tabelle 4-34 und Tabelle 4-35 geben einen Überblick über das Design der eingeschlossenen Studien bzw. der Intervention. Tabelle 4-36 zeigt die Charakteristika der Patienten der Studie 190-201. Die Interimanalyse der Folgestudie 190-202 umfasst alle Daten der Studie 190-201

sowie der nachfolgenden Studie 190-202 bis zum jeweiligen Datenschnitt; daher unterscheiden sich die Baseline-Charakteristika zwischen diesen Analysen nicht.

Tabelle 4-34: Charakterisierung der eingeschlossenen Studien

Studie	Studiendesign <RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	Population <relevante Charakteristika, z. B. Schweregrad>	Interventionen (Zahl der Patienten)	Studiendauer <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
190-201 [26]	Offene, einarmige, prospektive Interventionsstudie, historische Kontrolle nach 1:1 Paarbildung („matched“)	CLN2-Diagnose; genetisch gesichert und durch Analyse der TPP1-Aktivität bestätigt; frühes bis mittleres Krankheitsstadium (Score von 3-6 auf der kombinierten Motor-/Sprach-Skala (ML-Skala) und von mindestens 1 auf jeder Einzelskala), Krampfanfälle und deren Medikation aus Sicht des Prüfarztes stabil, zwischen drei und 15 Jahre alt, keine andere neurologische Erbkrankheit, keine andere, mit dem Verlust kognitiver Fähigkeiten verbundene, neurologische Erkrankung, keine frühere Stammzell-, Gentherapie oder ERT zur Behandlung der CLN2	Cerliponase alfa 300 mg ICV alle zwei Wochen (n=24)	Screening: ≤3 Tage vor der ICV-Implantation Chirurgische Implantation des ICV-Zugangs: Nicht länger als 14 Tage vor der ersten Infusion des Prüfpräparates Dosiseskulation (30 mg, 100 mg, 300 mg, Patienten Nr. 1-9): Jeweils mindestens vier Wochen pro Dosis, anschließend fixe Dosis (maximal tolerierte Dosis oder 300 mg alle zwei Wochen) über bis zu 48 Wochen Stabile Dosis (alle Patienten): Bis zu 48 Wochen Visite zu Studienende oder bei vorzeitigem Abbruch: ±3 Tage bezogen auf die letzte Infusion	4 Kliniken: Universität, Hamburg Kinderkrankenhaus; Bambino Gesù (Italien); Evelina Kinderkrankenhaus (Vereinigtes Königreich); Nationwide Children’s Hospital (USA) Durchführung: 13.09.2013 bis 30.11.2015	<u>Patientenrelevante, primäre Endpunkte:</u> Sicherheit und Verträglichkeit von Cerliponase alfa  Anteil der Patienten, die auf die Behandlung mit Cerliponase alfa nach 48 Wochen angesprochen hatten: Kein (Netto-)Abfall des ML-Scores um zwei Punkte oder Wert von Null in Woche 48  <u>Andere patientenrelevante, Endpunkte:</u> Verschiedene Analysen der adaptierten CLN2-Skala unter Behandlung mit Cerliponase alfa (siehe Tabelle 4-39)  Lebensqualität (explorativ nach SAP; PedsQL™ und CLN2-spezifisches QoL-Instrument)

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

Studie	Studiendesign	Population	Interventionen (Zahl der Patienten)	Studiendauer	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
	<RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	<relevante Charakteristika, z. B. Schweregrad>		<ggf. Run-in, Behandlung, Nachbeobachtung>		
				Nachbeobachtung (Sicherheit): 6 Monate ±1 Woche nach der letzten Infusion		
190-202 [27]	Offene, einarmige prospektive Inter- ventionsstudie, fortlaufende Folgestudie der Studie 190-201	Patienten, die die Studie 190-201 abgeschlossen hatten (48 Wochen)  Ausschluss von Patienten, deren Score auf der kombinierten Motor- /Sprach-Skala (ML- Skala) in der Studie 190-201 um $\geq 3$ Punkte abgefallen war  Ausschluss von Patienten mit einem Score = 0	Cerliponase alfa 300 mg ICV alle zwei Wochen (n=23)	Screening: Zwei Wochen beginnend in Woche 47 der Studie 190-201  Behandlungsphase: Beginnend mit der ersten Infusion in der Studie 190-202 (Baseline) über bis zu 239 Wochen  Nachbeobachtung: 6 Monate nach der letzten Infusion oder Aufnahme in eine durch BioMarin gesponserte Studie oder ein Register	4 Kliniken: Universität, Ham- burg, Kinderkranken- haus Bambino Gesù (Italien), Evelina Kinderkrankenhaus (Vereinigtes König- reich), Nationwide Children’s Hospital (USA)  Durchführung: 02.02.2015 (erster Patient in Studie 190- 202 eingeschlossen), fortlaufend	<u>Patientenrelevante primäre Endpunkte:</u>  Langzeitsicherheit von Cerliponase alfa  Anteil der Patienten, die auf die Behandlung mit Cerliponase alfa angesprochen hatten: Kein (Netto-)Abfall des ML- Scores um zwei Punkte oder Wert von Null in Woche 48  <u>Andere patientenrelevante, Endpunkte:</u>  Verschiedene Analysen der adaptierten CLN2- Skala unter Behandlung mit Cerliponase alfa  Lebensqualität unter Langzeitbehandlung mit Cerliponase alfa (explorativ nach SAP; PedsQL™, CLN2- spezifisches QoL-Instru- ment und EQ-5D)



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

<b>Studie</b>	<b>Studiendesign</b> <RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	<b>Population</b> <relevante Charakteristika, z. B. Schweregrad>	<b>Interventionen (Zahl der Patienten)</b>	<b>Studiendauer</b> <ggf. Run-in, Behandlung, Nachbeobachtung>	<b>Ort und Zeitraum der Durchführung</b>	<b>Primärer Endpunkt; patientenrelevante sekundäre Endpunkte</b>
190-901 Supplemental Report [56]	Patientenregister zum natürlichen Verlauf (historische Kontrolle der Studie 190-201), prospektive und retrospektive Erhebung	Patienten des Registers DEM-CHILD mit genetisch gesicherter CLN2; mindestens 2 HML-Scores im Abstand von mindestens 6 Monaten und im Alter von mindestens 36 Monaten; mindestens ein HML-Score $\geq 3$ Punkte und mindestens zwei Werte zwischen 1-5 Punkten	Keine Intervention 41 von insgesamt 74 Patienten erfüllten die Eignungskriterien	Nicht zutreffend (fortlaufendes Patientenregister).	2 Kliniken: Universität Hamburg, Deutschland (n=33), Universita degli Studi di Verona, Italien (n=8) Fortlaufend (teilweise retrospektiv)	Nicht zutreffend. Rate der Krankheitsprogression und deren Variabilität nach Auftreten klinischer Symptome unter Einsatz der HML-Skala Beschreibung der Krankheitscharakteristika
CLN2: klassische spätinfantile neuronale Ceroid-Lipofuszinose Typ 2; ICV: intracerebroventrikulär						

Tabelle 4-35: Charakterisierung der Interventionen – Weitere Studien mit dem zu bewertenden Arzneimittel [37]

Studie	Dosierung von Cerliponase alfa	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in- Phase etc.</i>
190-201	<p>Screening: <math>\leq 3</math> Tage vor der ICV-Operation</p> <p>Zur ICV-Verabreichung von Cerliponase alfa wurden den Patienten ein ICV-Reservoir und eine ICV-Kanüle chirurgisch implantiert. Die erste Dosis Cerliponase alfa konnte den Patienten während oder nach der Operation verabreicht werden, mindestens aber innerhalb von 14 Tagen nach der Operation.</p> <p>Dosiseskalation: Die ersten 9 Patienten nahmen an einer Dosiseskalation teil und wurden Kohorten aus jeweils drei Patienten zugeteilt. Diese Patienten erhielten initial Dosen von 30, 100 bzw. 300 mg Cerliponase alfa ICV alle zwei Wochen über jeweils mindestens vier Wochen (Kohorten 1, 2 bzw. 3). Die jeweils nächst höhere Dosis wurde nur verabreicht, wenn die Patienten die niedrigere Dosis vertragen hatten (keine unakzeptable Toxizität). Die Vertreter eines unabhängigen Datenkontrollkomitees (DMC) entschieden auf Basis der Sicherheitsdaten über die Dosiseskalationen. Die Patienten der Kohorten 1-3 wurden nacheinander nach Freigabe der jeweiligen Dosis durch das DMC eingeschlossen (siehe Abbildung 10).</p> <p>Stabile Phase: Nach Abschluss der Dosiseskalation, erhielten alle Patienten Cerliponase alfa in fixer Dosierung (300 mg oder die höchste, tolerierte Dosis) ICV alle zwei Wochen über bis zu 48 Wochen. Patienten, die nicht an der Dosiseskalation teilgenommen hatten, wurden direkt mit 300 mg Cerliponase alfa behandelt.</p> <p>Vor jeder Cerliponase alfa-Infusion wurden die Patienten stationär aufgenommen und während und nach den Infusionen engmaschig überwacht.</p> <p>Visite zu Studienende oder bei vorzeitigem Abbruch: <math>\pm 3</math> Tage nach der letzten Dosis</p> <p>Nachbeobachtung zur Sicherheit: 6 Monate nach der letzten Dosis (<math>\pm 1</math> Woche), vorausgesetzt die Patienten nahmen nicht an der Folgestudie 190-202 teil</p>	<p>Bei Auftreten schwerer Überempfindlichkeitsreaktionen wie Angioödem oder Stridor sollte die Infusion abgebrochen werden.</p> <p>Den Patienten wurde etwa 30 Minuten vor den Infusionen ein Anti-Histaminikum (und wenn angezeigt ein Anti-Pyretikum) in altersgerechter Dosierung verabreicht.</p> <p>Ein altersgerechtes Sedativum konnte im Ermessen der Prüfarzte vor den Infusionen eingesetzt werden (Amendment 2 des Studienprotokolls: 26.07.2013).</p> <p>Die Dosen begleitender Anti-Konvulsiva oder anderer Medikationen (Myoklonus, Tremor, Agitation oder Schmerzen) sollten während der Studie möglichst nicht verändert werden.</p>

<b>Studie</b>	<b>Dosierung von Cerliponase alfa</b>	<b><i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in- Phase etc.</i></b>
Studie 190-202 (Fortlaufende Folgestudie der Studie 190-201)	Den Patienten wird 300 mg Cerliponase alfa alle zwei Wochen ICV verabreicht. Die Behandlungsdauer beträgt bis zu 239 Wochen. Nachbeobachtung zur Sicherheit: 6 Monate nach der letzten Infusion, vorausgesetzt die Patienten nehmen nicht an einer anderen BioMarin-Studie oder einem Register teil.	Bei Auftreten schwerer Überempfindlichkeitsreaktionen wie Angioödem oder Stridor soll die Infusion abgebrochen werden. Die Prüferärzte können die Patienten nach eigenem Ermessen mit einer altersgerechten Dosis eines Anti-Histaminikums (und falls erforderlich eines Antipyretikums) vorbehandeln. Dem Alter entsprechende Sedativa konnten vor den Infusionen entsprechend der üblichen Praxis verabreicht werden.

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

Tabelle 4-36: Charakterisierung der Studienpopulationen – Weitere Untersuchungen mit dem zu bewertenden Arzneimittel und Untersuchungen zum natürlichen Verlauf

Baseline- Charakteristikum	190-201 [26] N=24	1:1 Paarbildung („matched“) <sup>1</sup>					190-202 [27] N=24
		190-201/202 ITT N=22	190-901 N=22	Hamburg N=33	Verona N=8	Gesamt N=41	
Mittleres Alter (SD) bei Studienein-schluss in Jahren	4,3 (1,24)	4,7 (0,93) (300 mg-Baseline)  Matching Kriterium ≤12 Monate Altersunterschied	4,7 (0,77)	n.a.	n.a.	n.a.	Wie 190-201
Mittleres Alter (SD) bei Diagnose-stellung	3,4 (1,07) Jahre	n.a.	n.a.	59,7 (18,97) Monate, ca. 5 Jahre (n=24)	59,9 (9,08) Monate, ca. 5 Jahre	59,8 (16,90) Monate, ca. 5 Jahre (n=32)	Wie 190-201
Mittleres Alter (SD) zum Zeitpunkt der ersten Symptome	n.a.	n.a.	n.a.	34,6 (8,60) Monate (n=24)	39,3 (9,60) Monate (n=8)	35,8 (8,94) Monate (n=32)	
Weiblich, n (%)	15 (63%)	13 (59%)	5 (23%)	11 (33%)	2 (25%)	13 (32%)	Wie 190-201
Männlich, n (%)	9 (38%)	9 (41%)	16 (73%)	20 (61%)	4 (50%)	24 (59%)	Wie 190-201
Geschlecht unbekannt, n (%)	0	0	1 (5%)	2 (6%)	2 (25%)	4 (10%)	
Abstammung, n (%)							Wie 190-201
Asiatisch	1 (4%)	n.a.	n.a.	n.a.	n.a.	n.a.	
Afro-amerikanisch	0	n.a.	n.a.	n.a.	n.a.	n.a.	
Weiß	23 (96%)	n.a.	n.a.	n.a.	n.a.	n.a.	
Andere	0	n.a.	n.a.	n.a.	n.a.	n.a.	
Ethnische Gruppe		n.a.	n.a.	n.a.	n.a.	n.a.	

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

Baseline- Charakteristikum	190-201 [26] N=24	1:1 Paarbildung („matched“)¹		190-901 [56] N=41 (auswertbare Patienten vor Paarbildung)			190-202 [27] N=24
		190-201/202 ITT N=22	190-901 N=22	Hamburg N=33	Verona N=8	Gesamt N=41	
Hispanisch oder Latino	1 (4%)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Nicht hispanisch oder Latino	23 (96%)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Genotyp, n (%)							
c.622C>T	5 (21%)	5 (23%)	7 (32%)	10 (30%)	1 (13%)	11 (27%)	Wie 190-201
c.509-1G>C	2 (8%)	2 (9%)	2 (9%)	2 (6%)	0	2 (5%)	
c.622C>T und c.509-1G>C	2 (8%)	2 (9%)	5 (23%)	8 (24%)	0	8 (20%)	
c.622C>T und andere	4 (17%)	4 (18%)	1 (5%)	2 (6%)	0	2 (5%)	
c.622C>T, fehlend		0	1 (5%)	0	3 (38%)	3 (7%)	
c.509-1G>C und andere	4 (17%)	3 (14%)	2 (9%)	6 (18%)	0	6 (15%)	
andere	7 (29%)	6 (27%)	4 (18%)	4 (12%)	2 (25%)	6 (15%)	
fehlend				1 (3%)	2 (25%)	3 (7%)	
Mittlerer (SD) ML-Score, [Minimum, Maximum]							Wie 190-201
Screening	3,7 (0,95) [3-6]		n.a.	n.a.	n.a.	n.a.	
Baseline	3,6 (1,06) [2-6]		n.a.	n.a.	n.a.	n.a.	

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

Baseline-Charakteristikum	190-201 [26]	1:1 Paarbildung („matched“)¹		190-901 [56]			190-202 [27]
	N=24			N=41 (auswertbare Patienten vor Paarbildung)			N=24
		190-201/202 ITT	190-901	Hamburg	Verona	Gesamt	
		N=22	N=22	N=33	N=8	N=41	
300 mg-Baseline	3,5 (1,20) [1-6]			n.a.	n.a.	n.a.	
ML/HML-Score	300 mg Baseline (n=23)	300 mg Baseline³		Bei Diagnose²	Bei Diagnose²	Bei Diagnose²	Wie 190-201
6	2 (9%)	2 (9%)	2 (9%)				
5	2 (9%)	2 (9%)	2 (9%)	4 (12%)	0	4 (10%)	
4	5 (22%)	5 (23%)	5 (23%)	7 (21%)	3 (38%)	10 (24%)	
3	11 (48%)	10 (45%)	10 (45%)	1 (3%)	3 (38%)	4 (10%)	
2	2 (9%)	2 (9%)	2 (9%)	6 (18%)	1 (13%)	7 (17%)	
1	1 (4%)	1 (5%)	1 (5%)	1 (3%)	0	1 (2%)	
0	0			3 (9%)	0	3 (7%)	
Fehlend				11 (33%)	1 (13%)	12 (29%)	
Mindestens eine Vorerkrankung, ≥15%/SOK und PT n (%)	24 (100%)						24 (100%)⁴
Kongenitale, familiäre, genetische Erkrankungen	24 (100%)			n.a.	n.a.	n.a.	24 (100%)
NCL	24 (100%)			n.a.	n.a.	n.a.	24 (100%)
Erkrankungen des Gastrointestinaltrakts	8 (33%)			n.a.	n.a.	n.a.	12 (50%)
Obstipation	<15%			n.a.	n.a.	n.a.	6 (25%)
Dysphagie	<15%			n.a.	n.a.	n.a.	5 (21%)

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

Baseline- Charakteristikum	190-201 [26] N=24	1:1 Paarbildung („matched“)¹			190-202 [27] N=24		
		190-201/202 ITT N=22	190-901 N=22	190-901 [56] N=41 (auswertbare Patienten vor Paarbildung)			
				Hamburg N=33	Verona N=8	Gesamt N=41	
Hypersalivation	<15%			n.a.	n.a.	n.a.	4 (17%)
Allgemeine Erkrankungen und Beschwerden am Verab- reichungsort	4 (17%)						12 (50%)
Entwicklungs- verzögerung	<15%			n.a.	n.a.	n.a.	5 (21%)
Gangstörung	<15%			n.a.	n.a.	n.a.	6 (25%)
Erkrankungen des Nerven- systems	23 (96%)						24 (100%)
Epilepsie	18 (75%)			n.a.	n.a.	n.a.	19 (79%)
Ataxie	14 (58%)			n.a.	n.a.	n.a.	15 (63%)
Sprachstörung	11 (46%)			n.a.	n.a.	n.a.	14 (58%)
Sprechstörung	<15%			n.a.	n.a.	n.a.	10 (42%)
Psychomoto- rische Fähig- keiten einge- schränkt	10 (42%)			n.a.	n.a.	n.a.	10 (42%)
Dysarthrie	6 (25%)			n.a.	n.a.	n.a.	6 (25%)
Myoklonus	<15%			n.a.	n.a.	n.a.	6 (25%)
Dystonie	0			n.a.	n.a.	n.a.	6 (25%)
Krampfanfall	5 (21%)			n.a.	n.a.	n.a.	6 (25%)

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

Baseline- Charakteristikum	190-201 [26] N=24	1:1 Paarbildung („matched“)¹			190-202 [27] N=24	
		190-201/202 ITT N=22	190-901 N=22	190-901 [56] N=41 (auswertbare Patienten vor Paarbildung)		
			Hamburg N=33	Verona N=8	Gesamt N=41	
Sprachstörung entwicklungs- bedingt	5 (21%)		n.a.	n.a.	n.a.	5 (21%)
Erniedrigter Muskeltonus	4 (17%)		n.a.	n.a.	n.a.	5 (21%)
Tremor	<15%		n.a.	n.a.	n.a.	4 (17%)
Symptome, die zur Diagnose führten, n (%)						
Krampfanfälle	n.a.		22 (67%)	8 (100%)	30 (73%)	
Ataxie	n.a.		18 (55%)	5 (63%)	23 (56%)	
Sprachschwierig- keiten	n.a.		17 (52%)	7 (8%)	24 (59%)	
Abnahme allgemeiner Fähigkeiten	n.a.		11 (33%)	6 (75%)	17 (41%)	
Andere motorische Probleme	n.a.		12 (36%)	1 (13%)	13 (32%)	
Myoklonus	n.a.		8 (24%)	3 (38%)	11 (27%)	
Verhaltensauf- fälligkeit	n.a.		5 (15%)	4 (50%)	9 (22%)	
Unbekannt	n.a.		9 (27%)	0	9 (22%)	
Verminderte feinmotorische Fähigkeiten	n.a.		3 (9%)	4 (50%)	7 (17%)	



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

Baseline- Charakteristikum	190-201 [26] N=24	1:1 Paarbildung („matched“) <sup>1</sup>			190-202 [27] N=24
		190-201/202 ITT N=22	190-901 N=22	190-901 [56] N=41 (auswertbare Patienten vor Paarbildung)	
			Hamburg N=33	Verona N=8	Gesamt N=41
Verlust der Sehkraft	n.a.		4 (12%)	3 (38%)	7 (17%)
Muskelschwäche	n.a.		1 (3%)	3 (38%)	4 (10%)
<sup>1</sup> 1:1 Paarbildung („matching“): Identischer ML-Skalenwert und Altersunterschied ≤12 Monate jeweils zu Baseline <sup>2</sup> HML-Score zum Diagnosezeitpunkt in der Studie 190-901 <sup>3</sup> Ein identischer ML/HML-Score wurde zur Paarbildung herangezogen. <sup>4</sup> Andauernde UEs/UEs, die bereits in der Studie 190-201 aufgetreten waren, wurden als Vorerkrankung dokumentiert. n.a.: not available (nicht verfügbar) CSR 190-201: Tabelle 11.2.1, Tabelle 11.2.2, Tabelle 11.2.3, Tabelle 14.1.7 CSR 190-202: Tabelle 11.2.2 Modul 2.7.3: Tabelle 2.7.3.3.2.1.1 [57] 190-901 Supplemental Report: Tabelle 8.2, Tabelle 8.3, Tabelle 8.4. Tabelle 8.5					

Tabelle 4-37: Baseline-Werte der auswertbaren Patientenpopulation (190-901) auf der HML-, HMLV- und HMLVS-Skala vor und nach 1:1 Paarbildung mit Patienten der Studie 190-201 [57]

Messinstrument Studie 190-901	Methode zur Bestimmung des Baselinewerts in Studie 190-901 <sup>1</sup>	Mittelwert (SD)	
		Studie 190-901 (auswertbare Patienten)	Studie 190-201/202 (ITT <sup>2</sup> )
<i>HML Skala / vor 1:1 Paarbildung</i>			
0-6-Punkte HML-Skala	Schätzung durch Interpolation, 901: mittlerer Wert	4,4 (0,78) (n=41)	3,5 (1,20) (n=23)
0-6-Punkte HML-Skala	Schätzung durch Interpolation, 901: Erster HML-Wert <6	4,4 (0,78) (n=41)	3,5 (1,20) (n=23)
0-6-Punkte HML-Skala	“Windowing”, 901: mittlerer Wert	4,4 (0,78) (n=41)	3,5 (1,20) (n=23)
0-6-Punkte HML-Skala	Schätzung durch Interpolation, 901: mittlerer Wert <6	4,4 (0,78) (n=41)	3,2 (0,94) (n=21) („Wirksamkeits“- Population)
<i>HML Skala / nach 1:1 Paarbildung</i>			
0-6-Punkte HML-Skala	Schätzung durch Interpolation, 901: mittlerer Wert, <u>1:1 Paarbildung</u> : ML/HML-Score	3,5 (1,19) (n=23)	3,5 (1,20) (n=23)
0-6-Punkte HML-Skala	Schätzung durch Interpolation, 901: mittlerer Wert, <u>1:1 Paarbildung</u> : ML/HML-Score und Alter (primäre Analysepopulation)	3,4 (1,09) (n=22)	3,4 (1,09) (n=22) <sup>3</sup>
0-6-Punkte HML-Skala	Schätzung durch Interpolation, 901: mittlerer Wert, <u>1:1 Paarbildung</u> : ML/HML-Score und Genotyp	3,5 (1,19) (n=23)	3,5 (1,20) (n=23)
<i>HMLV Skala und HMLVS Gesamtskala / vor 1:1 Paarbildung</i>			
0-9-Punkte HMLV-Skala	Schätzung durch Interpolation, 901: mittlerer Wert	7,2 (1,14) (n=39) <sup>4</sup>	6,3 (1,34) (n=23)
0-12-Punkte HMLVS (Gesamtskala)	Schätzung durch Interpolation, 901: mittlerer Wert	9,1 (1,79) (n=39) <sup>4</sup>	8,0 (1,83) (n=23)
<p><sup>1</sup> 190-201/202: Baseline = letzter Wert vor der ersten 300 mg-Infusion. 190-901: Baseline = mittlerer Wert für eine Reihe aufeinander folgender gleicher Skalenwerte &lt;6 (oder &lt;9 HMLV oder Gesamtscore &lt;12) sofern zutreffend, oder erster HML-Wert unter 6.</p> <p><sup>2</sup> Wenn nicht anders angegeben.</p> <p><sup>3</sup> Ein Patient aus der ITT Population der Studie 190-201 konnte nicht gepaart werden, da der Altersunterschied zu groß war (21 Monate)</p> <p><sup>4</sup> Bei zwei Patienten lagen keine Daten zum Sehvermögen vor.</p> <p>ISE: Tabelle 2.1.1, Tabelle 2.1.1.a, Tabelle 2.1.2, Tabelle 2.1.3, Tabelle 2.1.5, Tabelle 2.1.6, Tabelle 2.1.7, Tabelle 2.2.1</p>			

***Beschreibung des Designs der Studie 190-201: Studienprotokoll Amendment 5.0 vom 24.09.2014 [37], SAP Version 3.0 vom 25.03.2016 [51] und SAP ISE Version 3.0 vom 25.04.2016 [58]***

Das Design der Studie wurde gemäß TREND-Statement berichtet („Transparent Reporting of Evaluations with Non-randomized Designs“; siehe Tabelle 4-75). Das TREND-Statement stellt den internationalen Berichtsstandard für nicht randomisierte Studien dar (siehe Abschnitt 4.2.5.1) und wird für die Studie 190-201/202 als maßgeblich eingestuft. Das Design der Studie 190-901 (historische Kontrolle) ist in Tabelle 4-77 auf freiwilliger Basis gemäß STROBE-Statement (Strengthening the Reporting of Observational Studies in Epidemiology) unter Berücksichtigung des RECORD-Statements (REporting of studies Conducted using Observational Routinely-collected health Data) beschrieben.

Die Ergebnisse der Studie 190-201 werden in diesem Nutzendossier als maßgeblich betrachtet, da die Studie abgeschlossen ist und entsprechend SAP ausgewertet wurde. Deren Folgestudie 190-202 („extension“) dauert noch an, und finale Auswertungen entsprechend Studienprotokoll und SAP liegen noch nicht vor. Die Ergebnisse der (nicht geplanten) Zwischenanalysen der Folgestudie 190-202 werden aus rein informativen Gründen dargestellt z. B. um zu bewerten, ob der Behandlungseffekt dauerhaft ist, oder zur Analyse des Sicherheitsprofils von Cerliponase alfa oder des ICV-Ports unter Langzeittherapie. Die Hypothesentestung bezog sich auf eine Behandlungsdauer von 48 Wochen, und daher können derzeit nur die Ergebnisse der Studie 190-201 als beweisführend bewertet werden. Es liegen zwei Versionen des SAP vor, der originäre SAP sowie ein SAP, der auf Feedback der Zulassungsbehörden für die „Integrated Summary of Efficacy“ (ISE) erstellt wurde (im Folgenden als „ISE SAP“ bezeichnet). Der ursprüngliche SAP der Studie 190-201/202 sieht eine primäre Responder-Analyse (vs. festgesetzte, historische Kontrollrate) vor. Auf Empfehlung der Zulassungsbehörden wurde im ISE SAP eine Responder-Analyse nach 1:1 Paarbildung der Teilnehmer der Studie 190-201/202 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) als primärer Endpunkt festgelegt. Da die Vergleichbarkeit der Gruppen, insbesondere bei einer historischen Kontrolle, von entscheidender Bedeutung ist, werden die Ergebnisse der ISE-Analyse als für die Nutzenbewertung von Cerliponase alfa maßgeblich eingestuft. Aus Gründen der Transparenz und Vollständigkeit sowie, um internationale Berichtsstandards zu erfüllen (z. B. Darstellung des primären Endpunkts), werden die Hauptanalysen der Studie 190-201 ohne 1:1 Paarbildung zusätzlich berichtet. Da die Studie 190-202 noch nicht abgeschlossen ist und bislang nur Zwischenanalysen vorliegen, werden die Ergebnisse der Studie 190-202 lediglich ergänzend zusammen mit den Daten der Studie 190-201 dargestellt.

Bei der Studie 190-201 handelt es sich um eine offene, nicht kontrollierte, prospektive Dosisescalationsstudie der Phase 1/2. Die Studie hatte zum Ziel, die Sicherheit, Verträglichkeit, Pharmakokinetik und Wirksamkeit von Cerliponase alfa (ICV verabreicht) bei Patienten mit CLN2 zu untersuchen. Ein unabhängiges Datenkontrollkomitee (DMC, „Data Monitoring Committee“) überprüfte die klinischen Ergebnisse, Laborbefunde und Sicherheitsdaten während jeder der Eskalationsschritte sowie regelmäßig während der stabilen Dosierungsphase. Die Vertreter des DMCs, darunter mindestens ein Allergologe oder Immunologe und ein Biostatistiker, waren an der Studie nicht direkt beteiligt. Im Falle inakzeptabler,

arzneimittelbedingter Toxizität konnte BioMarin nach Rücksprache mit dem DMC die weitere Rekrutierung aussetzen, solange die abschließende Analyse der Daten noch ausstand. Kriterien zur Definition inakzeptabler, arzneimittelbedingter Toxizität waren *a priori* festgelegt.

Für die Studie 190-202 wurde kein DMC berufen.

### *Ziele*

Die Studie 190-201 hatte protokollgemäß folgende primäre Ziele:

- Sicherheit und Verträglichkeit von Cerliponase alfa, ICV verabreicht mittels eines implantierten Ports (Reservoir und Kanüle), bei Patienten mit CLN2
- Wirksamkeit von Cerliponase alfa nach 12 Monaten Behandlung im Vergleich zu unbehandelten CLN2-Patienten (basierend auf Literaturdaten zum natürlichen Verlauf), gemessen mittels einer CLN2-spezifischen Bewertungsskala

Die sekundären und explorativen Ziele der Studie sind in Tabelle 4-75 angegeben.

Die Studie 190-202 hatte folgende primäre Ziele:

- Langzeitsicherheit von Cerliponase alfa in einer Dosis von 300 mg alle 14 Tage bei Patienten mit CLN2
- Änderung der motorischen und sprachlichen Fähigkeiten, gemessen mittels einer leicht modifizierten Version der CLN2-spezifischen Bewertungsskala, bei Patienten mit CLN2 unter Behandlung mit Cerliponase alfa (300 mg alle 14 Tage).

Die sekundären und explorativen Ziele der Studie sind in Tabelle 4-76 angegeben.

Basierend auf intensiven Diskussionen mit den Zulassungsbehörden wurde die primäre Responder-Analyse im ISE SAP (verglichen mit dem ursprünglichen SAP) dahingehend geändert, dass den Patienten der Studie 190-201/202 jeweils ein vergleichbarer Patient der historischen Kontrollgruppe (Studie 190-901) zugeteilt wurde („1:1 matching“). Diese Änderung stimmt mit dem im Protokoll definierten Ziel der Studie 190-201 grundsätzlich überein, beschränkt dieses allerdings auf folgende Fragestellung:

- Wirksamkeit von Cerliponase alfa unter Einsatz einer CLN2-spezifischen Bewertungsskala im Vergleich zu unbehandelten Patienten (historische Kontrolle) [58].

### *Patientenpopulation:*

Voraussetzung für den Einschluss in die Studie war die Diagnose einer milden bis mittelschweren CLN2, definiert durch einen ML-Score von 3-6 Punkten und mindestens einem Punkt auf jeder der beiden Domänen (motorische Funktion/Sprache). Nach derzeitigem Kenntnisstand wird die Erkrankung bei unbehandelten Patienten mit einem ML-Score von 3-6 Punkten während der Dauer der Studie manifest, ihre neurologische Restfunktion reicht allerdings noch aus, um einen Behandlungserfolg ohne störende Bodeneffekte nachweisen zu

können. Es wurden Patienten in die Studie aufgenommen, deren Krampfeigung und anti-epileptische Begleitmedikation stabil waren (nach Einschätzung des Prüfarztes). Nur Patienten im Alter von mindestens drei und höchstens 15 Jahren wurden in die Studie eingeschlossen (für weitere Ein- und Ausschlusskriterien siehe Anhang 4-E). Nach Studien zum natürlichen Verlauf könnten Kinder unter drei Jahren während der Dauer der Studie klinisch stabil bleiben, was den Nachweis eines Behandlungseffekts erschwert hätte.

Die Diagnose CLN2 wurde durch Messung der TPP1-Enzymaktivität (Trockenblut, „dry blood spot“) bei Studieneinschluss sowie zu Baseline (Letzteres durch ein Zentrallabor) gesichert

#### *Fallzahlschätzung:*

Zur Fallzahlschätzung wurden Analysen aus einer validierten Datenbank mit Sitz in Hamburg (Deutschland) herangezogen, die Längsschnittdaten zum Krankheitsverlauf bei 30 unbehandelten Patienten mit genotypisch gesicherter CLN2 enthält. Demnach fällt der HML-Score nach Auftreten erster Symptome erwartungsgemäß innerhalb von 18 Monaten im Mittel um zwei Punkte bei einer SD von 1,8 ab. Unter der Annahme, dass der Score unter Behandlung bei vergleichbarer Variabilität nur um 0,5 Punkte innerhalb von 18 Monaten abnehme, wurde eine Fallzahl von 18 auswertbaren Patienten geschätzt, um bei einer Power von 90% und einem zweiseitigen Typ I-Fehler von 0,05 einen Unterschied nachweisen zu können. Bei einer geschätzten Abbruchquote von etwa 20%, sollten 22 Patienten mit CLN2 in die Studie aufgenommen werden. Die Zahl der Patienten wurde mit Amendement 3 des Studienprotokolls (vom 07.07.2014) von 22 in etwa 22 Patienten geändert. Während der Studie wurden zwei weitere, potenziell geeignete Patienten vorstellig und beide Patienten erfüllten die Eignungskriterien und wurden dementsprechend in die Studie aufgenommen (während der stabilen Dosisphase).

#### *Dosierung:*

Da Cerliponase alfa ICV zu verabreichen ist, wurden geeignete Patienten zur chirurgischen Implantation eines ICV-Ports (Reservoir und Kanüle) stationär eingewiesen, und die post-operative Erholungsphase war auf maximal 14 Tage beschränkt.

Insgesamt 22 Patienten waren geplant, und die Patienten wurden nach Erscheinen in die Studie aufgenommen. Die geplante Dosis von 300 mg wurde anhand von Tierstudien ermittelt. Die ersten neun Patienten nahmen an einer Dosisescalation teil und wurden nacheinander drei Dosisgruppen aus jeweils drei Patienten zugeteilt (Kohorte 1: 30 mg [erste drei Patienten], Kohorte 2: 100 mg und Kohorte 3: 300 mg). Während der Eskalationsphase sollte jeder der drei Patienten der einzelnen Dosiskohorten mindestens zweimal die jeweilige Dosis erhalten. Eine Dosisescalation (30 auf 100 mg bzw. 100 auf 300 mg) fand erst statt, nachdem Auswertungen bei der mit der jeweils niedrigeren Dosis behandelten Kohorte vorlagen. Nach Abschluss der Dosisescalation erhielten alle weiteren Patienten Cerliponase alfa in fixer Dosierung (300 mg oder die höchste Dosis, die die Patienten vertragen hatten) über mindestens 48 Wochen. Das DMC entschied in Absprache mit dem medizinischen Monitor von BioMarin und auf Basis der Sicherheitsdaten über die Dosisescalation. Das Studiendesign und die Dosierung sind in Abbildung 10 graphisch dargestellt.

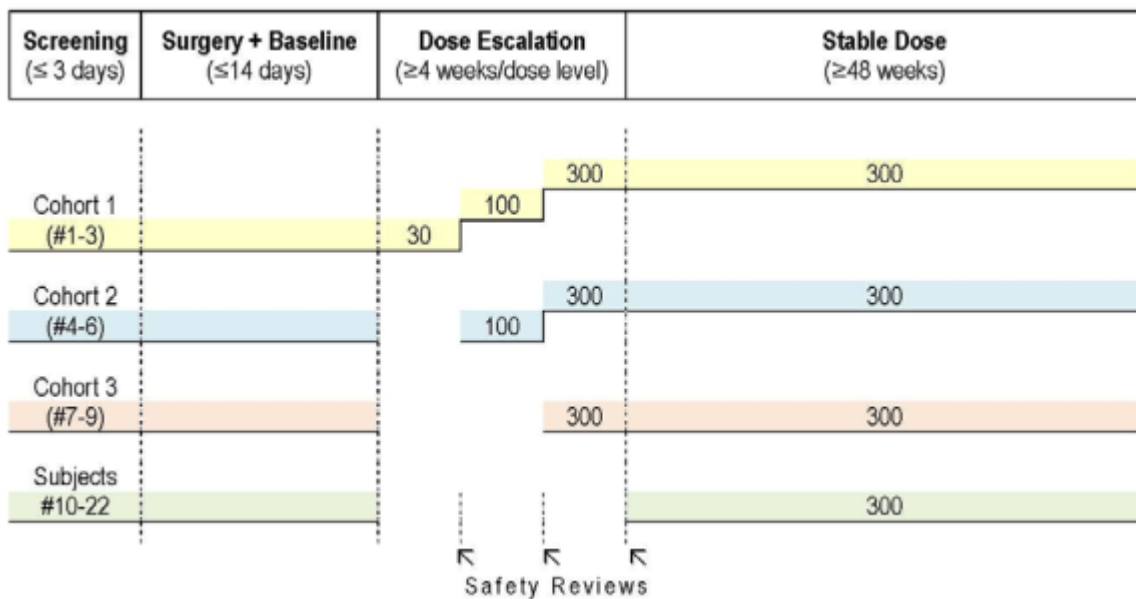


Abbildung 10: Studiendesign von Studie 190-201 und Dosierung von Cerliponase alfa während der Dosisescalation und der stabilen Dosierungsphase

Studienprotokoll 190-201: Figure 9.1.1 [37]

*Primärer und wichtige sekundäre Endpunkte:*

Der primäre Endpunkt der Studie war der Summenscore der ML-Skala. Jede Domäne des Endpunkts (motorische Funktion bzw. Sprache) wurde auf einer Skala von 0-3 bewertet, wobei ein Wert von drei bedeutet, dass der Patient fast nicht oder nicht eingeschränkt ist, und ein Wert von Null den vollständigen Funktionsverlust (zur Operationalisierung des Endpunkts siehe Tabelle 4-39). Eine Responder- und Ereigniszeitanalyse stellten die primären Auswertungen dar. Ein Ansprechen auf die Therapie war wie folgt definiert: Kein Abfall des ML-Scores um  $\geq 2$  Punkte, der bis zum Studienende erhalten bleibt, oder ML-Score = 0 in Woche 48 (Studientag 340 bezogen auf die erste Infusion von 300 mg). Mit der Primäranalyse (Responder- und Ereigniszeitanalysen) wurde Empfehlungen der Zulassungsbehörden entsprochen (ISE SAP Version 3.0 vom 25.04.2016).

Nach Diskussionen mit den Zulassungsbehörden wurde der SAP für die ISE geändert und eine Responder-Analyse nach 1:1 Paarbildung der Patienten der Studie 190-201/202 mit denjenigen der Studie 190-901 als Primäranalyse definiert. Ein Ansprechen auf die Therapie war wie folgt definiert: Abfall des ML-Scores um weniger als zwei Punkte in 48 Wochen. Diese Änderung erfolgte bereits vor den Wirksamkeitsanalysen [58]. Der ursprüngliche primäre Endpunkt („Slope“) blieb als wichtiger Endpunkt erhalten.

Das im Protokoll der Studie 190-201 definierte Erfolgskriterium (Abnahme des ML-Scores um zwei Punkte) basierte auf einer Kombination aus Expertenmeinungen und Literaturdaten, die einen durchschnittlichen Abfall des Messwerts um zwei Punkte innerhalb von 48 Wochen prognostizierten. Diese Annahme wird durch die Ergebnisse der nachfolgenden Studie 190-901

zum natürlichen Verlauf gestützt (Abfall von 2,1 Punkte pro 48 Wochen; Supplemental Report).

Die Hirnatrophie (z. B. totales Hirnvolumen, graue Substanz, Ventrikelvolumen) wurde mittels MRI als sekundärer Endpunkt erhoben, allerdings liegen bei den Patienten der historischen Kontrolle keine vollständigen MRI-Daten vor.

#### *Algorithmen zur Paarbildung („matching“) der Teilnehmer der Studie 190-201/202 und der Studie 190-901 (historische Kontrolle)*

1:1 Paarbildung, Methode #1 (primäre Analyse): Identischer ML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline; Dauer der Nachbeobachtung so lange wie möglich (mindestens sechs Monate bei Patienten der Studie 190-901), aber nicht länger als das vollständige Follow-up der gepaarten Patienten der Studie 190-201.

1:1 Paarbildung, Methode #2: Quadratsumme des Abstands zwischen ML-Score und Alter (in Monaten) zu Baseline ist  $\leq 36$  und mittleres Abstandsquadrat ist möglichst niedrig (Nachbeobachtung wie oben).

Many-to-one-Paarbildung: Paarbildung (jeweils zu Baseline) anhand der Variablen ML-Score, ML-Score und Alter und ML-Score und häufiger Genotyp (für weitere Details siehe Tabelle 4-75).

#### *Analysepopulationen*

Die ITT („Intention-to-treat“)-Population (190-201/202) setzte sich aus allen Studienteilnehmern zusammen, die Cerliponase alfa, unabhängig von der Dosis, erhalten hatten und bei denen zumindest ein Wirksamkeitsergebnis vorlag. Ein Patient brach die Studie nach nur einer Infusion ab, da er nicht bereit war, sich weiterhin den studienbedingten Maßnahmen zu unterziehen. Dieser Patient wurde aus allen Analysen einschließlich ITT ausgeschlossen, da der Patient nur eine Dosis während der Dosiseskaltionsphase erhalten hatte und der Abbruch nicht mit dem Behandlungsergebnis in Zusammenhang stand.

Ziel der primären Wirksamkeitsanalyse der ISE war es, die Wirksamkeit von Cerliponase alfa 300 mg alle zwei Wochen im Vergleich zur historischen Kontrolle anhand der Krankheitsprogression, gemessen mittels HML-Skala, zu bestimmen. Hierzu wurden den Patienten der ITT-Population der Studie 190-201/202 (Cerliponase alfa) jeweils auswertbare Patienten der historischen Kontrolle (190-901) zugeteilt („matching“). Die ITT-Population der Studie 190-201/202 entsprach der oben angegebenen Definition. Bei der auswertbaren Population der Studie 190-901 handelte es sich um Patienten, die die Haupteinschlusskriterien der Studie 190-201 erfüllten: Alter mindestens drei Jahre und mindestens ein HML-Score von mindestens drei Punkten im Alter von mindestens drei Jahren. Wenn nicht anders angegeben, beziehen sich die Analysen der Studie 190-901 auf Patienten, bei denen mindestens zwei HML-Werte im Bereich von 1-5 Punkten im Abstand von mindestens sechs Monaten vorlagen. Der primäre Algorithmus zur 1:1 Paarbildung zwischen den auswertbaren Patienten der Studie 190-901 und der ITT-Population der Studie 190-201/202 sah vor, dass nur Patienten, deren ML/HML-Score

zu Baseline exakt übereinstimmte und deren Altersunterschied so gering wie möglich war (allerdings nicht mehr als 12 Monate), einander zugeteilt wurden.

Ein zweiter 1:1 Algorithmus zur Paarbildung (Methode #2) wurde unterstützend eingesetzt. Dieser Algorithmus diente dazu, eine möglichst hohe Zahl an Paaren zwischen den Teilnehmern der Studien 190-201 und 190-901 zu erreichen, bei denen die Quadratsumme des Abstands zwischen ML-Score und Alter (in Monaten) zu Baseline  $\leq 36$  war und das mittlere Abstandsquadrat möglichst niedrig war.

In weiteren, unterstützenden Analysen wurde ein Many-to-one-Algorithmus zur Paarbildung genutzt; dieser basierte auf dem ML-Score oder ML-Score plus Alter oder ML-Score plus häufiger Genotyp (jeweils zu Baseline). Zusätzlich wurde die gesamte ITT-Population der Studie 190-201/202 mit den auswertbaren Patienten der Studie 190-901 ohne Paarbildung verglichen.

Sensitivitätsanalysen wurden bei folgenden Subpopulationen durchgeführt:

- Aus dieser Analyse wurden zwei der insgesamt 22 Teilnehmer ausgeschlossen, deren ML-Score während der gesamten Studie stabil bei sechs Punkten blieb (sogenannte „Wirksamkeits“-Population). Diesem Ausschluss lag die Rationale zugrunde, dass diese Patienten möglicherweise noch nicht die Phase der klinischen Progression erreicht hatten („decline phase“, die durch einen ML-Wert  $< 6$  Punkte gekennzeichnet ist).
- Ausschluss von Patienten mit einem stabilen ML-Score von sechs Punkten und einem ML-Wert unter drei Punkten zu Baseline. Patienten mit einem ML-Score  $< 3$  Punkte wurden aus Gründen der Konsistenz mit den Einschlusskriterien aus dieser Analyse ausgeschlossen (Alle Teilnehmer der Studie 190-201 hatten zur Screening-Visite einen ML-Score von  $\geq 3$  Punkte [Einschlusskriterium], allerdings nahm der ML-Score bei einigen Patienten bis zur Baseline-Erhebung auf  $< 3$  Punkte ab).

Die Per-Protocol (PP)-Population wurde zur Überprüfung der Sensitivität analysiert. Gemäß der Richtlinie ICH E9 war die PP-Population als alle Patienten der ITT-Population definiert, die den wesentlichen Protokollvorgaben entsprachen. Gründe für einen Ausschluss von Patienten waren vor der finalen Datenbanksperrung zu definieren. Nur Patienten, die wesentlichen Ein- und Ausschlusskriterien entsprachen, wurden in der PP-Population eingeschlossen: Einschlusskriterien: Alter  $\geq 3$  bis  $< 16$  Jahre; Diagnose der spätinfantilen CLN2, bestätigt durch TPP1-Enzymaktivität; milde bis mittelschwere Erkrankung, bestimmt mittels ML-Skala; stabile Krampfneigung; Ausschlusskriterien: Andere, erblich bedingte, neurologische Erkrankungen; frühere neurologische Erkrankungen, die mit kognitiven Beeinträchtigungen einhergehen; keine Vortherapie mit Stammzellen, Gentherapie oder ERT zur Behandlung der CLN2; keine anderen Prüfpräparate innerhalb von 30 Tagen vor und während der Studie). Die Patienten wurden mindestens 23 Wochen mit einer Dosis von 300 mg behandelt und erhielten 80% oder mehr der geplanten Infusionen während der 300 mg-Dosisphase.

Die Analyse zur Sicherheit umfasste alle Patienten, denen ein ICV-Port implantiert worden war.



Die Ergebnisse wurden getrennt für folgende Gruppen dargestellt: Kohorten 1, 2 und 3 während der Dosisescalation; stabile Dosisphase (SDO, „stable dose only“; 300 mg) und alle Patienten. Im Rahmen der ISE (Paarbildung mit historischen Kontrollen) wurden keine Analysen zur Sicherheit durchgeführt, da keine historischen Kontrolldaten vorliegen.

Die Definition von Baseline variierte kontextabhängig:

- Zur Überprüfung der Wirksamkeit von Cerliponase alfa in einer Dosis von 300 mg entsprach Baseline der letzten Beobachtung vor der ersten 300 mg-Infusion, entweder während der Dosisescalation oder der stabilen Dosisphase (im Weiteren „300 mg-Baseline“ genannt; CLN2-Scores und MRI). Diese wurde bei der primären Responder-Analyse zum Vergleich von Cerliponase alfa mit der historischen Kontrolle („matched“) eingesetzt (ISE). Bei den Analysen zur QoL und Denver II galt Baseline grundsätzlich als die letzte Beobachtung vor der ersten 300 mg-Infusion.
- Zur Überprüfung der Wirksamkeit von Cerliponase alfa in jeder Dosierung entsprach Baseline der letzten Beobachtung vor der ersten Infusion von Cerliponase alfa (im Weiteren „Studien-Baseline“ genannt); diese Definition galt auch bei den Analysen zur Sicherheit von Cerliponase alfa.
- Bei den Analysen zur Sicherheit, die auch den ICV-Port umfassten, galt Baseline als die letzte Beobachtung vor Implantation des ICV-Zugangs. Lag kein Wert vor der Implantation vor, entsprach Baseline der ersten Beobachtung nach der ICV-Implantation.

Die primären und sekundären Endpunkte (modifizierte ML-Skala und MRI) wurden während der Dosisescalation erhoben, die explorativen Endpunkte (QoL und Denver) allerdings nicht. Protokollgemäß stellte der Beginn der stabilen Dosisphase eine Art Neustart im Studienablauf dar, und unmittelbar vor der ersten Infusion der stabilen Dosisphase waren Untersuchungen (einschließlich ML-Score) geplant. Bei den Patienten, die an der Dosisescalation teilgenommen hatten, entsprach der Zeitpunkt der ersten 300 mg-Infusion nicht demjenigen der ersten Dosis von 300 mg während der stabilen Dosisphase. Daher wurde für diese Patienten ein Zeitfenster („window“) definiert. Dieser Algorithmus sah vor, dass zu jedem protokollgemäß geplanten Untersuchungszeitpunkt während der stabilen Dosisphase diejenige Beobachtung herangezogen wird (hierbei galt die erste 300 mg-Infusion als Ausgangspunkt), die der protokollgemäß Geplanten zeitlich am nächsten lag. Da die modifizierte ML-Skala und die MRI-Scans auch während der Dosisescalation erhoben wurden, haben diese Variationen im Ablaufplan auf diese einen weit geringeren Einfluss als auf die explorativen Endpunkte. Bei der primären Wirksamkeitsanalyse sowie Analyse zur Sicherheit wurden keine Zeitfenster eingesetzt.

Bei Patienten der Studie 190-901 (historische Kontrolle):

Bei den Analysen nach 1:1 Paarbildung gilt das Alter (in Monaten zum Zeitpunkt der Untersuchung), zu dem die Paarbildung stattfand, als Baseline.

Bei den Analysen, bei denen Many-to-one-Algorithmen zur Paarbildung eingesetzt wurden, wurden zwei verschiedene Baseline-Definitionen verwendet:

- Zunächst wurden diejenigen aneinandergrenzenden Visiten identifiziert, bei denen der HML-Score dem ML-Baselinewert des zugeteilten Patienten der Studie 190-201/202 entsprach. Dies konnte eine einzelne Visite sein oder auch eine Reihe aus mehreren Visiten. Lagen mehrere Visiten vor, galt der *mittlere Wert zwischen der ersten und der letzten dieser Visiten* als Ausgangspunkt (Baseline); oder
- Ähnlich wie oben, allerdings galt bei dieser Definition nicht der mittlere Wert zwischen der ersten und letzten aus einer Reihe aneinandergrenzender Visiten als Ausgangspunkt (Baseline), sondern der *Zeitpunkt der ersten Visite*.

Bei den Analysen ohne Paarbildung wurden ebenfalls zwei unterschiedliche Baseline-Definitionen eingesetzt:

- Identifikation der ersten Visite (d. h. das jüngste Alter), bei der der HML-Score <6 war, sowie einer Reihe aneinandergrenzender Visiten, bei denen dieser HML-Score erhalten blieb. Dies konnte eine einzelne Visite sein oder auch mehrere Visiten. War nur eine Visite vorhanden, entsprach diese Visite Baseline. Lagen mehrere Visiten vor, galt der *mittlere Wert zwischen der ersten und letzten dieser Visiten* als Ausgangspunkt (Baseline); oder
- Ähnlich wie oben, allerdings galt bei dieser Definition nicht der mittlere Wert zwischen der ersten und letzten aus einer Reihe aneinandergrenzender Visiten als Ausgangspunkt (Baseline), sondern der *Zeitpunkt der ersten Visite*.

#### *Dokumentierte Baseline-Charakteristika*

Folgende Baseline-Charakteristika wurden in der Studie 190-201 dokumentiert: Demographische Merkmale (Alter, Geschlecht und ethnische Gruppe), Körpergröße und Gewicht, Krankheitsmerkmale zu Baseline (z. B. Alter bei Diagnosestellung, Genotyp sowie CLN2-Scores zur Screening-Visite, vor Einleitung des Prüfpräparates und vor der ersten Infusion des Prüfpräparates in einer Dosis von 300 mg) und Krankheitsgeschichte (mittels MedDRA kodiert) [26]. Da die Zwischenanalysen der Folgestudie 190-202 den vollständigen Datensatz der Studie 190-201 einschließen, sind die Baseline-Charakteristika der Patienten der Studie 190-202 die gleichen. Aufgrund der teilweise retrospektiven Datenerhebung in der Studie zum natürlichen Verlauf (190-901) wurden die Charakteristika der Studienpopulation nicht systematisch erfasst [56].

#### *Subgruppen-Analysen*

In den Studienprotokollen der Studien 190-201 und 190-202 waren keine Subgruppen-Analysen festgelegt [37,45].

Bereits mit Amendement 2 des Protokolls vom 07.03.2013 (gültig zum Zeitpunkt, als der erste Patient eingeschlossen wurde) wurde festgelegt, dass der SAP gilt, sollte es Diskrepanzen

zwischen der Beschreibung der statistischen Analyse im Protokoll und dem finalen SAP geben. Gemäß SAP (Version 3.0), der sowohl für die Studie 190-201 als auch 190-202 gilt, konnten folgende Subgruppen-Merkmale hinsichtlich einer möglichen Effektmodifikation (Wirksamkeit) untersucht werden: häufiger Genotyp (insbesonder c.622C>T- und c.509-1G>C-Mutationen), demographische Merkmale, Studienzentrum oder geographische Region. Darüber hinaus konnten verschiedene Behandlungsintervalle untersucht werden (z. B. gesamtes Dosisintervall, nach Einleitung der 300 mg-Dosis, während der stabilen Dosisphase oder nach einer bestimmten Wirkstoffexposition, wie 90 Tage oder 1.000 mg des Prüfpräparates). Alle Subgruppen-Analysen waren explorativ und spezifische statistische Methoden waren nicht festgelegt [51]. Homogenitäts- oder Interaktionstests waren nicht vorgesehen.

Im ISE SAP der Studien 190-201/202 sind ebenfalls verschiedene Behandlungsintervalle angegeben, die denjenigen des SAP entsprechen [58].

#### *Studiendauer*

Studie 190-201 [26]:

Studienbeginn (d. h. Einverständniserklärung des ersten Patienten) war der 13. September 2013 und die hier berichteten Ergebnisse bilden den Datenschnitt vom 30. November 2015 ab. Alle Patienten der Studie 190-201 haben 48 Wochen der Behandlung abgeschlossen. Die mittlere (SD) Behandlungsdauer bei einer Dosis von 300 mg betrug 47,1 (10,90) Wochen und die gesamte Behandlungsdauer (jede Dosis) 49,5 (13,23) Wochen. Das ursprüngliche Protokoll (vom 7. März 2013) wurde den Zulassungsbehörden im Mai 2013 zugeleitet. Im Verlauf der Studie gab es insgesamt fünf Protokoll-Amendements. Amendment 1 wurde den Zulassungsbehörden im Mai 2013 vorgelegt. Der erste Patient wurde in die Studie aufgenommen, als Amendment 2 gültig war. Die Amendments 3-5 wurden den Zentren gleichzeitig zur Verfügung gestellt. Die Zentren arbeiteten bis zu dem Zeitpunkt unter Amendment 2, als Amendment 5 (vom 24. September 2014) freigegeben wurde.

Studie 190-202: Datenschnitt vom 15. Oktober 2015 [27]:

Bis zu diesem Datenschnitt wurden 18 Teilnehmer der Studie 190-201 in die Folgestudie aufgenommen und über mehr als ein Jahr mit Cerliponase alfa behandelt (bei diesen 18 Patienten, die vor dem angegebenen Datenschnitt in die Studie 190-202 aufgenommen worden waren, bis zu 31 zusätzliche Wochen, verglichen mit der Studie 190-201). Die letzte Patienten-Visite findet voraussichtlich am 15. Dezember 2020 statt.

Historischer Vergleich („Integrated summary of Efficacy“, ISE): Datenschnitt vom 15 Oktober 2015

Zum Datenschnitt vom 15. Oktober 2015 waren lediglich 9 Teilnehmer der Studie 190-201 über weitere 24 Wochen im Rahmen der Folgestudie 190-202 behandelt worden; daher wurden zur primären, vergleichenden Wirksamkeitsanalyse die Daten nach 48 Wochen der Studie 190-201 herangezogen. Dies diente dazu, sicher zu stellen, dass alle Patienten gleich lange mit dem Prüfpräparat behandelt wurden, und die den Analysen zugrunde liegende Datenbasis aussagekräftig ist. Bei der Analyse der Responder-Raten nach Kaplan-Meier und der Slope-

Analyse wurde demgegenüber die gesamte Studiendauer (190-201 und 190-202) berücksichtigt.

Studie 190-202: Datenschnitt vom 3. Juni 2016 [34].

Auf Ersuchen der FDA fand im Juni 2016 eine nicht geplante Analyse der Studienergebnisse statt. Die Behandlungsdauer (300 mg-Dosisintervall) betrug bei 7 Patienten (29%) >74-84 Wochen, bei fünf Patienten (21%) >84-96 Wochen und bei 11 Patienten (46%) >96 Wochen. Dementsprechend waren bis zu diesem Zeitpunkt 96% der Patienten über mehr als 72 Wochen behandelt worden.

Studie 190-202: Datenschnitt vom 1. November 2016 [59–62]

Ebenfalls auf Ersuchen der FDA wurden mit Datenschnitt zum 1. November 2016 weitere Follow-up-Analysen durchgeführt. Bis zu diesem Zeitpunkt waren alle Patienten über mindestens 96 Wochen mit Cerliponase alfa in fester Dosierung behandelt worden („stable dose period“).

#### *Datenschnitte der eingeschlossenen Analysen*

Im September 2014 und Januar 2015 wurden Zwischenanalysen der Studie 190-201 durchgeführt. Wie geplant und vorab festgelegt (DMC-Charta), fanden während der Dosiseskulation, jeweils nachdem alle drei Patienten einer Kohorte mindestens zwei Infusionen in der jeweiligen Dosierung (30, 100 oder 300 mg) erhalten hatten, Treffen des DMC statt. Bei diesen Treffen wurde bestimmt, ob die Sicherheitsdaten ausreichen, um die nächst höhere Dosisstufe einzuleiten, und/oder die verbleibenden Patienten der stabilen Dosisphase zugeteilt werden. Der Abschlussbericht der Studie 190-201 liegt vor, und alle Patienten der ITT-Population hatten bis zum Berichtszeitpunkt 48 Wochen der Behandlung abgeschlossen. In diesem Dossier werden die ISE-Analyse (historische Kontrolle mit Paarbildung; diese wird als maßgeblich für die Bewertung des Zusatznutzens von Cerliponase alfa eingestuft) sowie die Wirksamkeitsanalysen aus dem Abschlussbericht der Studie 190-201 dargestellt.

Der Abschlussbericht der Studie 190-202 liegt noch nicht vor. Zwischenanalysen der Studien 190-201/202 zur Sicherheit und Verträglichkeit fanden am 15. Oktober 2015 statt. Auf Ersuchen der FDA wurden die Daten der Studie 190-201/202 zum Datenschnitt Juni 2016 ausgewertet. Die jüngsten Wirksamkeitsanalysen, d. h. die Analysen auf Anfrage der FDA mit Datenschnitt im November 2016, werden zur zusätzlichen Information dargestellt

Die Ergebnisse des Abschlussberichts der Studie 190-201 und die Ergebnisse zum Datenschnitt im Juni 2016 werden zur Sicherheit/Verträglichkeit von Cerliponase alfa dargestellt.

#### *Analysepopulationen und Patientenfluss*

Studie 190-201 [26]:

Ursprünglich war der Einschluss von 22 Patienten geplant, mit Amendment 3 (7. Juli 2014) wurde die Fallzahl allerdings von 22 auf „etwa 22“ Patienten geändert. In zwei der

Studienzentren wurden etwa zur gleichen Zeit zwei weitere mögliche Kandidaten identifiziert, und diese eigneten sich zum Einschluss in die Studie.

Insgesamt 24 Patienten wurden in die Studie eingeschlossen. Jeweils drei Patienten nahmen in Kohorte 1 und 2 an der Dosisescalation teil und vier Patienten in Kohorte 3. Ein Patient der Kohorte 3 brach die Studie nach nur einer Dosis des Prüfpräparates ab und wurde durch einen anderen Patienten ersetzt. Vierzehn Patienten begannen direkt mit der stabilen Dosisphase. Die verbleibenden 23 Patienten schlossen die stabile Dosisphase der Studie 190-201 planmäßig ab (24 Infusionen). Alle 23 Patienten wurden anschließend in die Folgestudie 190-202 aufgenommen.

Alle Patienten wurden in die ITT-Population aufgenommen und schlossen die Behandlung ab mit einer Ausnahme (n=23/24). Dieser Patient brach die Studie auf Ersuchen der Eltern nach nur einer Infusion ab, da er nicht in der Lage war, weiterhin an den studienbedingten Maßnahmen teilzunehmen. Da der klinische Review keine Hinweise auf bedeutende Protokollverletzungen lieferte, waren die ITT- und PP-Population identisch.

Zwei Patienten, deren ML-Score von Baseline bis zum Ende von Woche 48 unverändert bei sechs Punkten blieb (Maximum, d. h. kein Abfall), wurden aus der sogenannten „Wirksamkeits“-Population ausgeschlossen (n=21, Sensitivitätsanalyse). Diese Patienten befanden sich vermutlich noch nicht in der Phase der klinischen Progression („decline phase“). Eine Untergruppe dieser Population (n=18) umfasste nur Patienten, deren Ausgangs-ML-Score (300 mg-Baseline) drei, vier oder fünf Punkte betrug.

Zusammenfassend wurden folgende Populationen analysiert:

- Alle eingeschlossenen Patienten: n=24
- ITT-Population (alle Patienten mit Ausnahme eines Patienten, der nur eine Infusion erhalten hatte): n=23 (primäre Analysepopulation)
- PP-Population (ITT ohne bedeutende Protokollverletzungen): n=23 (identisch mit ITT)
- „Wirksamkeits“-Population (ITT, ausgenommen zwei Patienten, deren ML-Score von sechs Punkten während der Studie unverändert blieb): n=21
- „Efficacy“-Population („Wirksamkeits“-Population wie oben, aber beschränkt auf Patienten mit einem ML-Score von drei, vier oder fünf Punkten zur 300 mg-Baseline): n=18
- Zur Sicherheit analysierte Population (alle Patienten, denen ein ICV-Port implantiert worden war): n=24

Diese Analysepopulationen blieben bei den Follow-up-Analysen der Folgestudie 190-202 unverändert (Datenschnitte vom 15.10.2015 und Juni 2016).

Studie 190-202:

Zum Datenschnitt der ersten Zwischenanalyse (15. Oktober 2015) hatten fünf von insgesamt 23 Patienten die Studie 190-201 noch nicht abgeschlossen und bei der Analyse wurden keine Erhebungen aus der Studie 190-202 berücksichtigt. Diese fünf Patienten wurden in die Studie 190-202 eingeschlossen, nachdem sie die Studie 190-201 planmäßig beendet hatten. Nur die ersten 9 Patienten waren im Rahmen der Studie 190-202 über weitere 24 Wochen behandelt worden. Zum Zeitpunkt einer zusätzlichen Zwischenanalyse im Juni 2016 auf Anforderung der FDA nahmen alle 23 Patienten an der Studie teil.

Studie 190-201 vs. historische Kontrolle („matched“; ISE) [57]:

Die Studie 190-202 stellt eine Erweiterung („extension“) der Studie 190-201 dar, und alle Patienten, die an der fortlaufenden Studie 190-202 teilnehmen, haben die 48-wöchige stabile Dosisphase der Studie 190-201 abgeschlossen. Daher wurde auf separate Auswertungen oder eine Zusammenfassung der Wirksamkeitsdaten der beiden Studien verzichtet. Die historische Kontrollgruppe (n=74) wurde auf eine „auswertbare“ Population aus 41 Patienten beschränkt, indem Patienten, die bestimmte Eignungskriterien der Studie 190-201 nicht erfüllten, aus dieser ausgeschlossen wurden (siehe oben „Natürlicher Verlauf“).

Eine Responder- und Ereigniszeitanalyse stellten die Primärauswertungen der ISE dar. Bei diesen Analysen wurde den Patienten der ITT-Population der Studie 190-201/202 jeweils 1:1 Patienten der historischen Kontrolle zugeteilt, vorausgesetzt, der Baseline-ML/HML-Wert stimmte exakt überein und der Altersunterschied (Baseline) war so gering wie möglich (betrug aber nicht mehr als 12 Monate; siehe oben für andere Paarbildungskriterien). Diese Kriterien trafen auf 22/23 Patienten (ITT) der Studie 190-201/202 zu; bei einem Patienten war eine Paarbildung nicht möglich, da der Altersunterschied zu groß war (21 Monate). Entsprechend umfassten die 1:1 „Matching“-Analysen insgesamt 22 Patienten der ITT-Population.

*Charakterisierung der Patientenpopulation zu Studienbeginn und/oder Baseline*

Studie 190-201 [26]:

Die Behandlungsgruppe bestand aus 9 (38%) männlichen und 15 (63%) weiblichen Patienten im mittleren Alter ( $\pm$  SD) von 4,3 (1,24) Jahren (Bereich: 3,0 bis 8,0 Jahre; Median 4,0 Jahre). Davon waren die Mehrzahl Weiße (n=23, 96%) und nur ein Patient (4%) war asiatischer Herkunft. Die Patienten erkrankten durchschnittlich im Alter von 3,4 (SD 1,07) Jahren (Median 3,0 Jahre) und die Patienten waren zwischen 2,5 und 6,3 Jahre alt, als die Erkrankung ausbrach (CSR 190-201: Tabelle 11.2.2). Bei 17 Patienten (71%) wurde entweder eines oder beide der häufigsten Allele (c.622C>T und c.509-1G>C) nachgewiesen und andere Mutationen bei 7 Patienten (29%; Tabelle 4-36).

Insgesamt 20/24 Patienten hatten zur Screening-Visite einen ML-Score von 3-4 Punkten (Mittelwert  $\pm$  3,7 $\pm$ 0,95) und entsprechend einen ML-Score im unteren (d. h. schlechteren), protokollgemäßen Bereich (Einschlusskriterium: 3-6 Punkte). Zur Baseline-Visite betrug der mittlere (SD) ML-Score 3,6 (1,06) Punkte. Zur Screening-Visite lagen alle ML-Scores

zwischen drei und sechs Punkten, sodass das Einschlusskriterium bei allen Patienten erfüllt war. Zwei Patienten wurden mit einem ML-Score von sechs in die Studie aufgenommen, d. h. diese befanden sich wahrscheinlich noch in einem frühen Stadium der Progression. Bei zwei Patienten nahm der ML-Score zwischen der Screening- und Baseline-Visite auf zwei Punkte ab. Diese Patienten waren dennoch für die Studie geeignet, da sich das entsprechende Einschlusskriterium protokollgemäß auf den Screening-Wert bezog. Bevor die Patienten erstmalig eine Dosis von 300 mg erhielten (300 mg-Baseline, d. h. unmittelbar vor der ersten Infusion von 300 mg), betrug der ML-Score 3,5 Punkte (Mittelwert, SD 1,20). Zu diesem Zeitpunkt (300 mg-Baseline) hatten drei Patienten einen ML-Score unter drei (n=1 ML-Score von 1; n=2 ML-Score von 2). Dementsprechend war die Krankheitslast, ausgedrückt als ML-Score, bei den meisten Patienten bereits zu Studienbeginn (Screening) erheblich. Vor der ersten Infusion von 300 mg (300 mg-Baseline) hatten 61% der Patienten (n=14/23) einen ML-Score von drei oder niedriger und nur 17% (n=4/23) einen ML-Score von 5-6 (Tabelle 4-36).

Alle Patienten hatten mindestens eine Vorerkrankung, wovon die meisten die SOCs „Kongenitale, familiäre und genetische Erkrankungen“ (100%) und „Erkrankungen des Nervensystems“ (96%) betrafen. Die meisten Vorerkrankungen (PT) des Nervensystems waren Epilepsie (75%), Ataxie (58%), Sprachstörung (46%), und psychomotorische Fähigkeiten eingeschränkt (42%).

Alle Patienten der Studie 190-201 (n=24, 100%) hatten vor der ersten Cerliponase alfa-Infusion bereits mindestens eine Vorbehandlung erhalten, meistens Propofol (88%), Midazolam (67%), Paracetamol (50%), Cefuroxim (46%), Sufentanil (46%), Levetiracetam (42%) und Natriumvalproat (42%). Einige dieser häufig dokumentierten Vorbehandlungen (z. B. Propofol) wurden möglicherweise nur im Rahmen der chirurgischen Implantation des ICV-Ports verabreicht, d. h. die Patienten wendeten diese vor Studienbeginn nicht regelmäßig an [55].

Alle Patienten erhielten Begleitmedikationen und/oder Nahrungsergänzungsmittel. Die häufigsten Begleitmedikationen waren Paracetamol (92%), Propofol (79%), Midazolam (75%), Natriumvalproat (75%), Diazepam (71%), Cetirizin (67%), Ibuprofen (67%), Levetiracetam (50%), Prednisolon (50%), Sevofluran (46%) und Valproinsäure (46%) (CSR 190-201: Tabelle 14.1.8.2).

*Responder-Analyse nach 1:1 Paarbildung (ISE): Studien 190-201 und 190-901 (historische Kontrolle) [57]*

Tabelle 4-36 zeigt die Baseline-Charakteristika der Cerliponase alfa-Gruppe und der historischen Kontrolle nach 1:1 Paarbildung. Insgesamt 22 der 23 Patienten der ITT-Population (190-201) und 22 der „auswertbaren“ Patienten der historischen Kontrollgruppe (190-901) erfüllten die Kriterien zur 1:1 Paarbildung. Der mittlere (SD) Altersunterschied zwischen der Cerliponase alfa-Gruppe und der historischen Kontrolle betrug zum Zeitpunkt der Paarbildung 1,5 (2,87) Monate. Bei 13 der Paare (59%) stimmte das Alter (in Monaten) exakt überein und bei nur vier Patienten war der Altersunterschied größer als zwei Monate. Diese Daten belegen die Güte der Paarbildung (ISE: Tabelle 0.1).

Insgesamt 16 der 22 Patienten (73%) der historischen Kontrollgruppe (190-901) und 9 der 22 Patienten (41%) der Studie 190-201/202 waren männlich. Daher war die Geschlechterverteilung zwischen den Gruppen nicht ausgeglichen. Allerdings stellt das Geschlecht keinen Prognosefaktor dar und hat keinen Einfluss darauf, wie schnell die Erkrankung voranschreitet. Wäre das Geschlecht als zusätzliches Kriterium zur Paarbildung eingesetzt worden, hätte dies die Qualität der Paarbildung eingeschränkt, die auf prognostisch bedeutenderen Faktoren (Alter und ML/HML-Score zu Baseline) beruht.

Hinsichtlich des Genotyps wurden bei insgesamt 14 der 22 Patienten (64%) der Studie 190-901 verglichen mit 9/22 Patienten der Studie 190-201/202 (41%) die beiden häufig auftretenden Allele (c.622C>T und/oder c.509-1G>C) nachgewiesen. Entsprechend wiesen mehr Patienten der Studie 190-201/202 als der Studie 190-901 („matched“) zwei der eher selten auftretenden Allele auf (6 Patienten [28%] 190-201/202, 4 Patienten [18%] 190-901). Der Behandlungsunterschied (Änderung des ML-Scores) zugunsten von Cerliponase alfa war allerdings vergleichbar, unabhängig davon, ob zur Paarbildung (jeweils zu Baseline) der ML-Score plus Art der CLN2-Mutation, der ML-Score allein oder der ML-Score plus Alter herangezogen wurde (ISE: Tabelle 4.1.9).

Die Gruppe bestehend aus allen auswertbaren Patienten der Studie 190-901 (historische Kontrolle) hatte zu Baseline einen um etwa einen Punkt höheren (besseren) ML-Score als die Patienten der Studie 190-201/202, unabhängig von der eingesetzten Methodik. Hinsichtlich des MLV- und Gesamt-Scores zeigte sich ein ähnlicher Trend (Tabelle 4-37). Bei den Analysen ohne Paarbildung galt bei den Patienten der Studie 190-901 der erste HML-Score <6 Punkte (d. h. das jüngste Alter) als Baseline. Dieser Unterschied ist wahrscheinlich eine Folge der Baseline-Definition, die zum Ziel hatte, die Phase der Progression („decline phase“) zu erfassen.

Bei den meisten Patienten der Studie 190-201 als auch der historischen Kontrolle (betrifft die Gruppe aller auswertbaren Patienten) deuten die vorliegenden Daten auf einen typischen, spätinfantilen Phänotyp hin. Kinder werden üblicherweise dann bei NCL-Spezialisten vorgestellt, wenn Ataxien oder zunehmend nicht anders erklärbare Krampfanfälle auftreten, die Sprachentwicklung verzögert ist und/oder die Kinder in ihrer Entwicklung stehen bleiben [9]. Ataxie (58%), Epilepsie (75%), Krampfanfälle (21%) und Sprachprobleme (42%) wurden bei den Teilnehmern der Studie 190-201/202 sehr häufig als Vorerkrankungen dokumentiert. Bei den Patienten des DEM-CHILD-Programms führten Krampfanfälle (73%), Ataxie (56%), Sprachprobleme (59%) sowie ein genereller Abbau körperlicher Funktionen am häufigsten zu der Diagnose (Tabelle 4-36). In einer Untersuchung bei italienischen Kindern mit NCL zeigten alle Kinder mit genetisch gesicherten CLN2-Mutationen (n=29) einen spätinfantilen Phänotyp [63].

Bislang gibt es keine krankheitsmodifizierende oder kurative Behandlung irgendeiner Form der NCL. Zudem sprechen CLN2-Patienten häufig nicht auf die zur unterstützenden Therapie eingesetzten Anti-Epileptika an [10]. Nach einer Umfrage in Deutschland und UK unter 19 Familien mit Kindern mit CLN2, wird den Kindern eine Vielzahl an Medikationen zur symptomatischen Therapie verschrieben, u. a. bei Krampfanfällen, übermäßiger Sekretbildung,



Muskelzucken/Dystonie, Stimmungsschwankungen, Schlafstörungen sowie Problemen, die die Mobilität, das Sehen und die Kommunikation betreffen. Weitere Details sind in dem Konferenzbericht nicht angegeben [53]. Dies stimmt mit der Studie 190-201/202 überein, bei der die Zahl der Begleitmedikationen ebenfalls hoch war. Informationen zu früheren Therapien oder Begleitmedikationen liegen bei der historischen Kontrollgruppe nicht vor.

### ***Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext***

Die Ein- und Ausschlusskriterien der Studie 190-201 waren nicht restriktiv. Die Einschränkungen hatten hauptsächlich zum Ziel, den Behandlungseffekt auf den Verlust motorischer und sprachlicher Fähigkeiten während der progredienten Phase der Erkrankung („decline phase“) zuverlässig erfassen zu können.

Sowohl die Patienten der Studie 190-201 als auch der Studie 190-901 (historische Kontrolle) stellen repräsentative Stichproben der in Deutschland zu erwartenden Patientenpopulation dar. In der klinischen Studie waren zwölf Patienten aus Deutschland, drei aus dem Vereinigten Königreich, sechs aus Italien und drei aus den USA. Das durchschnittliche Alter bei Diagnosestellung lag bei den Patienten beider Studien in einem Bereich, der nach bisheriger Kenntnis des Krankheitsverlaufs bei CLN2-Patienten zu erwarten ist. In beide Studien wurden Patienten in einer frühen bis mittleren Phase der Progression eingeschlossen (gekennzeichnet durch einen ML-Score über zwei Punkte), d. h. Patienten, die sich wahrscheinlich auch außerhalb einer Studie in Behandlung begeben hätten. Die Diagnose wird oftmals erst spät gestellt [9]; dies wird durch die Studien gestützt, deren Teilnehmer bereits bei Studieneinschluss relativ deutlich beeinträchtigt waren. Insgesamt 61% der 41 auswertbaren Patienten der Studie zum natürlichen Verlauf (190-901 Supplemental Report) hatten einen HML-Score von  $\geq 2$  Punkten (44%  $> 2$  Punkte in Einklang mit dem Einschlusskriterium der Studie 190-201), allerdings fehlten die Angaben bei 12 Patienten (29%; Supplemental Report 190-901: Tabelle: 14.15) [56].

Die Mehrzahl der Patienten der Studie 190-201 war bei Krankheitsbeginn zwischen drei und  $< 5$  Jahre alt (50%) und das Mindest- bzw. Höchstalter betrug 2,5 bzw. 6,3 Jahre. Bei Studieneinschluss waren die Patienten 3,0-8,0 Jahre alt (CSR 190-201: Tabelle 11.2.1, Tabelle 11.2.2). Die Anamnese der Patienten war hinsichtlich der Art der Vorerkrankungen wie Ataxie, Epilepsie und Sprachstörungen typisch (Tabelle 4-36). Bei Patienten mit NCL beginnt die Diagnosestellung üblicherweise in dem Alter, in dem die Patienten symptomatisch werden, d. h. bei jungen Kindern, deren Entwicklung stehen bleibt oder rückläufig ist und/oder bei denen schwere Epilepsien unbekannter Ursache auftreten. Bei diesen Patienten sollte nach üblichen klinischen Standards die TPP1- (und PPT1)-Aktivität gemessen werden [10]. Ein Konsortium international ausgewiesener Experten, darunter Experten aus Deutschland, hat einen Algorithmus zur Diagnose der CLN2 entwickelt. Bei Patienten, die mit typischen Symptomen vorstellig werden (neu aufgetretene, nicht anders erklärbare Krampfanfälle und/oder Ataxien, denen Störungen der frühen Sprachentwicklung und/oder der Entwicklung insgesamt vorausgegangen waren) sollte die Verdachtsdiagnose CLN2 gestellt und weitere diagnostische Maßnahmen eingeleitet werden [9]. In die Studie von Steinfeld et al. (Studienzentrum in Hamburg) wurden die Patienten auf Grundlage klinisch neurodegenerativer Symptome

eingeschlossen, die in der frühen Kindheit auftraten, wie schwer behandelbare Krampfanfälle, fortschreitender Verlust motorischer, sprachlicher sowie geistiger Fähigkeiten und später einsetzender Sehverlust. Zudem zeigten sich elektronenmikroskopisch Zeichen, die auf eine spätinfantile NCL hinwiesen [5]. Abgesehen von den sehr wenigen Kindern, die aufgrund ihrer familiären Vorgeschichte frühzeitig auf CLN2 getestet werden, wird die Erkrankung auf Basis der typischen, klinischen Manifestationen der Erkrankung sowie des Alters der Patienten diagnostiziert. Wie die Studie von Steinfeld et al. zeigt [5], scheint dies auch in Deutschland übliche klinische Praxis zu sein.

Während der Studie 190-201 betrug die Exposition gegenüber Cerliponase alfa bei asiatischen Patienten 1,3 und bei Weißen 21,6 Patientenjahre (94% der Gesamtexposition von 22,9 Patientenjahren; CSR 190-201: Tabelle 14.3.6.1.12). Per Datenschnitt im Juni 2016 war die Exposition auf 42 Patientenjahre bei weißen und 2,5 Patientenjahre bei asiatischen Patienten angestiegen (Datenschnitt Juni 2016: Tabelle 14.3.6.1.12). Ein Patientenjahr war hierbei als 52 Wochen der Exposition definiert. Zwölf der 24 Teilnehmer der Studie 190-201 (50%) wurden in Deutschland (Hamburg) rekrutiert. Die anderen Studienzentren befanden sich in Italien, UK und den USA mit wahrscheinlich insgesamt vergleichbaren Gesundheitssystemen.

Im Juni 2015 nahmen 24 medizinische Fachkräfte und Patientenvertreter weltweit an einer Online-Umfrage teil. Nach Auskunft der Experten werden zur symptomatischen Behandlung der CLN2 üblicherweise zahlreiche Medikationen eingesetzt, z. B. bei Krampfanfällen, Myoklonus, Spastik, Dystonie, Sekretbildung und Schmerzen [40]. Dies entspricht der Studie 190-201/202. Bei den Patienten der historischen Kontrollgruppe liegen keine Informationen zur Art der Behandlung oder Versorgung vor. Da es während der letzten Jahrzehnte keine bedeutenden Innovationen zur Behandlung der CLN2 gab, erscheint es unwahrscheinlich, dass sich die medizinische Versorgung zwischen den Patienten der historischen Kontrolle (die Mehrzahl wurde zwischen 1990 und 2009 rekrutiert; nur n=6/41 [14,6%] vor 1990) und den Teilnehmern der Studie 190-201/202 (erster Patient im September 2013 eingeschlossen) gravierend unterscheidet. In diesem Zusammenhang ist ebenfalls von Bedeutung, dass die Mehrzahl der historischen Kontrollen (n=33/41, 80,5%) und die Hälfte der Teilnehmer der Studie 190-201/202 in dem Zentrum in Hamburg rekrutiert wurden.

Die Rate, mit der die Erkrankung bei den Patienten der historischen Kontrolle (190-901, auswertbare Patienten) mit zunehmendem Alter voranschritt, war derjenigen bei einer anderen, unabhängigen und großen Patientenkohorte aus den USA, Kanada, UK und Australien sehr ähnlich (siehe z. B. Abbildung 8). Insgesamt deutet dies stark darauf hin, dass die historische Kontrollgruppe aus dem DEM-CHILD-Register eine repräsentative Stichprobe von Patienten mit CLN2 darstellt und der Therapierealität in Deutschland entspricht.

#### **4.4.2.3 Ergebnisse aus weiteren Untersuchungen**

In Tabelle 4-38 sind die Endpunkte, die in den eingeschlossenen Studien erhoben wurden, dargestellt.

Tabelle 4-38: Matrix der Endpunkte in den eingeschlossenen Studien mit dem zu bewertenden Arzneimittel

Studie	Mortalität	Morbidität	Gesundheitsbezogene Lebensqualität	Sicherheit und Verträglichkeit
190-201	Nein	Krankheitsaktivität gemessen mittels CLN2-spezifischer Instrumente	PedsQL™ CLN2-spezifischer Fragebogen	Gesamtzahl der UEs Gesamtzahl der UEs pro SOK Häufige UEs SUEs und Todesfälle UEs vom CTCAE Grad $\geq 3$ Studien- oder Behandlungsabbruch aufgrund eines UEs UEs besonderen Interesses (AEOSI): Überempfindlichkeits- und anaphylaktische Reaktionen (MedDRA SMQs) Infusionsassoziierte UEs (TRE)
190-202 (noch laufend)	Nein	Krankheitsaktivität gemessen mittels CLN2-spezifischer Instrumente	PedsQL™ CLN2-spezifischer Fragebogen EQ-5D <sup>1</sup>	Gesamtzahl der UEs Gesamtzahl der UEs pro SOK Häufige UEs SUEs und Todesfälle UEs vom CTCAE Grad $\geq 3$ Studien- oder Behandlungsabbruch aufgrund eines UEs UEs besonderen Interesses (AEOSI): Überempfindlichkeits- und anaphylaktische Reaktionen (MedDRA SMQs) Infusionsassoziierte UEs (TRE)
190-901 (historische Kontrolle)	Nein	Krankheitsaktivität gemessen mittels CLN2-spezifischer Instrumente	Nein	Nein

<sup>1</sup> Der Ergebnisse des EQ-5D-Messinstruments werden nicht dargestellt, da die Studie nicht abgeschlossen ist und Aussagen zum Verlauf unter Langzeittherapie nicht möglich sind.

#### 4.4.2.3.1 Krankheitsprogression auf CLN2-spezifischen Bewertungsskalen

Tabelle 4-39: Operationalisierung des Endpunkts „Krankheitsprogression auf CLN2-spezifischen Bewertungsskalen

Studie	Operationalisierung
190-201	<p>Die Krankheitsaktivität wurde mittels zweier CLN2-spezifischer Bewertungsskalen ermittelt: der modifizierten Hamburg-Skala und der WCMC-Skala. Beide Skalen bestehen aus vier Domänen (modifizierte Hamburg Skala: motorische und Sprachkompetenz, Sehvermögen und Krampfanfälle; WCMC: Gang, Sprache, Myoklonus und Ernährung) und werden durch den Kliniker bewertet. Jede Domäne beider Skalen deckt Werte zwischen 0 und 3 ab, sodass sich ein Summenscore von 0 (sehr schwer beeinträchtigt) bis zu 12 (weitgehend normal) ergibt. Die modifizierte Hamburg- bzw. WCMC-Skala wurden u.a. in zwei Studien zum natürlichen Verlauf erhoben (Steinfeld et al. 2002 [5] bzw. Worgall et al. 2007 [6]). Die originale Hamburg-Skala wurde in der Longitudinalstudie 190-901 zum natürlichen Verlauf der Erkrankung eingesetzt [39,56].</p> <p>Mit dem Ziel, objektive Bezugspunkte („anchors“) für die Studien 190-201/202 zu generieren, wurde in Kollaboration mit den Autoren der Hamburg-Skala eine geringfügig modifizierte Version der motorischen und der Sprachdomäne erarbeitet. Die Bezugspunkte erlauben eine zwischen den Studienzentren und über die Zeit einheitliche Bewertung und sind in Tabelle 4-29 dargestellt. Als primärer Endpunkt der Studie war der Summenscore aus der adaptierten, motorischen und Sprachdomäne (ML-Skala) definiert (Scores von 6 = weitgehend normal bis 0 = hochgradig beeinträchtigt).</p> <p><b>Maßnahmen zur Standardisierung der Bewertungen</b></p> <p>Die Methodik der Tests inklusive Training und Überwachung der Studienzentren war in einem Manual detailliert festgelegt: “Procedures for Training, Scoring, Videotaping and Reliability Testing of The Late Infantile Neuronal Ceroid Lipofuscinosis (LINCL) Clinical Scales Used for the Cerliponase alfa Program Studies” (datiert 29 April 2014) [37]. Darüber hinaus liegt ein Dossier zur Entwicklung und Validierung der ML-Skala vor: “A Brief Clinician-Reported Outcome Evidence Dossier Prepared For: The 0-6 Hamburg Scale in Patients with Neuronal Ceroid Lipofuscinosis Type 2 (CLN2). BioMarin Studies 190-901 and 190-201“ BioMarin (Version 2.0 datiert 16.05. 2016) [37].</p> <p>Alle Bewerter hatten vor Beginn der Studie ein Training zu absolvieren, um eine möglichst standardisierte Erhebung und Bewertung der Ergebnisse in Übereinstimmung mit dem Manual während der Studie zu gewährleisten. Professor Dr. Alfried Kohlschütter MD, ein ausgewiesener Experte in diesem Bereich und Entwickler der HML-Skala, führte die Trainings durch. Das Training umfasste einen Vortrag gefolgt von Videos von CLN2-Patienten, die anschließend gemeinsam diskutiert wurden. Die Erhebungen wurden in den Studienzentren regelmäßig überwacht, um sicher zu stellen, dass die Tests im Verlauf der Studie einheitlich erhoben werden. Wenn immer möglich sollte ein Prüfarzt die Tests durchgängig ausführen [37].</p> <p>Alle Tests wurden mittels Videokamera aufgezeichnet (geändert mit Amendment 3 des Studienprotokolls, 07.07.2014, ursprünglich alle 24 Wochen). Die Videoaufnahmen waren in Übereinstimmung mit dem Manual “Ratings Assessment Guidelines” standardisiert. Die Videoaufzeichnungen wurden nicht zur Bewertung herangezogen, sondern ausschließlich die klinischen Beobachtungen während der Tests [37].</p> <p>Zusätzlich zu den klinischen Bewertungen durch die Prüfarzte wurden die Eltern/Betreuer befragt. Während der Interviews wurden die Angehörigen/Betreuer gebeten, Angaben zur Gehfähigkeit, der Sprache, dem Sehvermögen, anomalen unwillkürlichen Bewegungen und der Ernährung der Kinder in der vorangegangenen Woche zu machen. Im Verlaufe der Interviews stellten die Prüfarzte zu jeder Fähigkeit gezielte Fragen. Die Interviews wurden nicht zur Erhebung des ML-Scores herangezogen, sondern dienten dazu, eventuelle Unstimmigkeiten verglichen mit den Beobachtungen des Prüfarztes aufzudecken oder wie der Patient am besten motiviert werden kann. Die Eltern und/oder Betreuer der Patienten sollten möglichst während des gesamten Tests und nicht nur während der Interviews anwesend sein. Die Tests sollten in der Muttersprache des Kindes abgehalten werden. Falls der Prüfarzt die Muttersprache des Patienten nicht beherrschte, stand ein zertifizierter Übersetzer bereit, der bewertete, ob die Sprache des Patienten verständlich war. Auch für die Eltern- /Betreuerinterviews konnte ein Übersetzer herangezogen werden [24,37].</p> <p>Die Tests sollten möglichst zu demselben Zeitpunkt während der Visiten stattfinden, vorzugsweise am Morgen vor anderen Untersuchungen und/oder den Infusionen. Falls der Patient in</p>

Studie	Operationalisierung
	<p>irgendeiner Weise beeinträchtigt war z.B. durch Sedation/Anästhesie oder Krankheit, sollten die Tests verschoben werden [24,37].</p> <p><b>Überwachungszeitpunkte der CLN2-Bewertungskalen</b></p> <p>Die Tests fanden zur Screening-Visite, vor der ersten Infusion (Baseline), an Tag vier nach der ersten Infusion, in Woche eins (Tag 4) nach jeder Dosisescalation, alle vier Wochen während jeder Dosisescalation, zu Beginn und alle 8 Wochen während der stabilen Dosisphase sowie bei Studienende bzw. vorzeitigem Abbruch statt [37]. Die Domäne zur Krampfneigung der Patienten wurde alle 12 Wochen während der stabilen Dosisphase erhoben.</p> <p><b>Praktische Ausführung</b> [24]</p> <p>Hinsichtlich der Einzeldomänen (Gang, Sprache und Sehvermögen) sollte der Prüfarzt versuchen, die bestmögliche Leistung des Patienten zu erreichen. Die Eltern/Betreuer und/oder Prüfarzte konnten die Patienten durch entsprechende Kommandos oder Aufforderungen motivieren.</p> <p><b>Domäne motorische Fähigkeiten/Gangbild</b> [24]</p> <p>Der Patient wurde gebeten, sich im Rahmen seiner/ihrer Möglichkeiten mit oder ohne Hilfsmittel fortzubewegen z.B. durch den Raum zu gehen oder vom Tisch zum Stuhl und zurück (etc.). Hilfsmittel umfassten in diesem Zusammenhang jede Form der Unterstützung (wie Hilfsgeräte, Wände, Möbel, Eltern/Betreuer), die zur Fortbewegung eingesetzt wurden. Beobachtet wurden alle/alle spontanen Bewegungen, das Gangbild/Ataxien, Stürze, die Fähigkeit, aus dem Sitzen aufzustehen sowie der Einsatz externer Hilfsmittel.</p> <p>Während der Interviews wurden die Eltern/Betreuer gebeten, Angaben zur Fortbewegung der Kinder während der letzten Woche zu machen. Mindestens folgende Fragen wurden dabei gestellt: Kann das Kind mit Hilfestellung gehen? Stürze/Verletzungen? Benötigt das Kind regelmäßig einen Rollstuhl/Buggy/Kinderwagen und wenn ja für wie lange? Kann das Kind krabbeln oder schnell wegrutschen?</p> <p><b>Sprachdomäne</b> [24]</p> <p>Der Prüfarzt bewertete das Sprachvermögen des Patienten einschließlich verständlicher Wörter und grammatikalisch korrekter Sätze. Die Qualität der Äußerungen wurde in die Bewertung einbezogen. Erfasst wurde beispielsweise, ob der/die Patient(in) in der Lage ist, Kommandos zu befolgen, sich spontan zu äußern und nachvollziehbare oder zielgerichtete Sätze (z.B. der Wunsch nach einem Spielzeug) zu formulieren. Bevorzugt wurden hierzu dem Patienten vertraute Gegenstände eingesetzt, um das Sprachvermögen zu überprüfen.</p> <p>Während der Interviews der Eltern/Betreuer wurden mindestens folgende Fragen gestellt, die sich auf die letzte Woche bezogen: Ist das Kind in der Lage, auf irgendeine Weise zu sprechen? Wie viele Wörter umfasst der Wortschatz des Kindes? Setzt das Kind zielgerichtete Sätze ein? Zu welchem Anteil ist die Sprache des Kindes für einen Fremden unverständlich?</p> <p><b>Sehvermögen</b> [24]</p> <p>Das Sehvermögen der Patienten wurde bewertet, vorzugsweise unter Einsatz nicht beweglicher Gegenstände. Falls der Patient nur auf sich bewegende Dinge reagierte, sollte dies dokumentiert werden. Der Prüfarzt bewertete, inwieweit der Patient in der Lage war, einen gewünschten Gegenstand oder eine Person zu erkennen und nach diesem/dieser zu greifen. Hierbei wurde jeweils der Abstand zwischen dem Patienten und dem Ziel erfasst (weit oder nah). Falls der Patient dies nicht konnte, wurde überprüft, ob er/sie auf ein helles Licht reagiert und in welchem Abstand.</p> <p>Die Eltern/Betreuer wurden zusätzlich befragt, ob das Kind in der letzten Woche, ihm/ihr bekannte Dinge oder Personen erkannt hat und sich bewegende Gegenstände verfolgen konnte.</p> <p><b>Myoklonus-Domäne (WCMC)</b> [24]</p> <p>Diese Domäne erfasst unwillkürliche Bewegungen des Patienten. Jede spontane und unwillkürliche anomale Bewegung einschließlich, aber nicht beschränkt auf Myoklonus, Dystonie, Chorea oder Athetose wurde beobachtet und die Reflexe inklusive Babinski wurden überprüft. Zusätzlich wurden die Eltern befragt, ob sich das Kind in der letzten Woche ungewöhnlich oder seltsam bewegt hat und dieses zu beschreiben.</p>

Studie	Operationalisierung
	<p><u>Ernährung (WCMC) [24]</u></p> <p>Die Domäne erfasst die Fähigkeit des Patienten, selbständig zu essen oder zu trinken. Falls das Kind bereits mittels Gastrostomieschlauchs ernährt wurde, entfiel diese Domäne. Bewertet wurde, ob das Kind ohne Aspiration schlucken konnte und welche Art Nahrung aufgenommen werden konnte. Zusätzlich wurde das Kind beim Trinken eines dünnflüssigen Getränks beobachtet.</p> <p>Während der Eltern-/Betreuerinterviews wurden gezielt Fragen hinsichtlich der Fähigkeiten des Kindes während der letzten Woche gestellt.</p> <p><u>Krampfanfälle [37]</u></p> <p>Die Zahl der Krampfanfälle und eventuelle Änderungen seit der letzten Erhebung wurden mittels Interview erfasst. Die Domäne wurde alle 12 Wochen erhoben, sodass sich die Bewertung auf die vorausgegangenen 12 Wochen bezog. Änderungen der Begleitmedikation waren jeweils zu dokumentieren. Die Eltern/Betreuer machten Angaben zur Krampfneigung des Patienten seit der letzten Erhebung: Kein Krampfanfall (Grand mal), 1-2 Krampfanfälle, 3 Krampfanfälle oder mehr als 3 Krampfanfälle (&gt;1 pro Woche).</p> <p><b>Analyse des primären Endpunkts der Studie 190-201 (ohne historische Kontrolle)</b></p> <p>Der primäre Endpunkt war der Anteil Patienten, die auf die Therapie angesprochen hatten (Responder; ITT). Dieser war definiert als Anteil der Patienten, bei denen der ML-Score (Hamburg) bis Woche 48 (Studientag 340 bezogen auf die erste 300 mg-Infusion) nicht (Netto) um zwei Punkte abgefallen war, oder in Woche 48 ein Wert von Null registriert wurde. Ein Nettoabfall (stabile Abnahme) um zwei Punkte entsprach einem Abfall des Scores um zwei Punkte oder mehr, der im Verlauf der Studie nicht wieder anstieg (z.B. Abfall um zwei Punkte und anschließend Anstieg um einen Punkt oder mehr). Die statistischen Analysen und Hypothesen sind in Tabelle 4-75 dargestellt. Die Rate der Krankheitsprogression über 48 Wochen gemessen mittels ML-Skala stellte eine bedeutende sekundäre Analyse dar.</p> <p>Die folgenden, weiteren Analysen wurden prospektiv berichtet [26]:</p> <ul style="list-style-type: none"> <li>- 6-Punkt ML-Skala</li> <li>- 9-Punkt MLV-Skala (ML plus Sehvermögen)</li> <li>- 12-Punkt MLVS Skala (MLV plus generalisierte epileptische Anfälle)</li> </ul> <p><u>Imputation</u></p> <p>Hinsichtlich der Slope-Analysen spielten fehlende Werte keine nennenswerte Rolle, da in den Analysen lediglich Vorher-/Nachher-Werte berücksichtigt wurden.</p> <p><b>Analyse der Ansprechrate und des Abfalls auf der ML/HML-Skala: Vergleich der Studie 190-201/202 mit der Studie 190-901 (historische Kontrolle mit Paarbildung, "matched")</b></p> <p>Nach Diskussionen mit den Behörden wurde die Primäranalyse der ISE in eine Responder-Analyse (Abfall des ML-Scores um &lt;2 Punkte/48 Wochen) mit 1:1 Paarbildung der Teilnehmer der Studien 190-201/202 und 190-901 (natürlicher Verlauf) geändert (ITT 190-201/202, auswertbare Patienten: 190-901). Zusätzlich wurde der Anteil der Patienten, deren ML-Score nicht Netto um zwei Punkte abgefallen war oder Patienten mit einem ML-Score = 0 zur letzten Erhebung, analysiert. Die Slope-Analyse blieb als bedeutender Endpunkt erhalten. Die motorische und Sprachdomäne sowie der MLV- und Gesamtscore wurden zusätzlich ausgewertet.</p> <p><u>Fehlende Werte</u></p> <p>Folgende Änderung des Datensatzes der Studie 190-901 (historische Kontrolle) wurden vor der Analyse vorgenommen: Alle imputierten Werte wurden identifiziert und gelöscht. Bei Visiten ohne Angabe einer Untersuchungsnummer wurde der Wert der letzten Visite eingegeben und bei Visiten, bei denen ein Alter von 0 Monaten eingegeben worden war, wurde angenommen, dass es sich um imputierte Werte zum Zeitpunkt der Geburt handelt. Die Daten des zweiten</p>

Studie	Operationalisierung
	<p>Zwillings zweier Zwillinge wurden gelöscht. Aus Gründen der Konsistenz mit den Ein- und Ausschlusskriterien der Studie 190-201/202 wurden alle Visiten bei Patienten unter drei Jahren und anschließend alle Visiten mit einem ML-Score von zwei oder niedriger gelöscht. Nur Patienten mit mindestens zwei ML-Score-Werten zwischen eins und fünf im Abstand von mindestens sechs Monaten wurden in die Analyse eingeschlossen.</p> <p>In der Studie 190-901 fanden keine regulären Erhebungen der ML-Skala statt und die Erhebungszeitpunkte unterschieden sich in der Studie 190-201 zwischen der Dosiseskulation und Patienten, die direkt mit 300 mg behandelt wurden. Um eine Analyse beider Patientengruppen zu bestimmten Zeitpunkten zu ermöglichen, wurden die ML-Ergebnisse der Patienten beider Studien zu diesen Zeitpunkten imputiert. Bei Patienten der Studie 190-901, bei denen sich der Score zwischen zwei Visiten nicht änderte, wurde dieser Wert monatlich eingegeben. Die nominalen Zeitpunkte waren Woche 25, 37, 49 und 73 (entsprechend Tag 169, 253, 337, und 505; Woche 73: Studie 190-202). Multiple Methoden der Imputation wurden eingesetzt:</p> <ul style="list-style-type: none"> <li>- Lineare Interpolation: Für den Fall, dass keine Werte vor und nach dem Analysetag vorlagen, aber ein Einzelwert innerhalb von 28 Tagen einschließlich, wurde der Einzelwert eingegeben.</li> <li>- "Windowing": Es wurde derjenige Wert innerhalb des Zeitfensters eingegeben, der dem Analysetag am nächsten lag. Lagen zwei Werte in demselben Zeitabstand vor, wurde jeweils der frühere Wert berücksichtigt.</li> </ul> <p>Weder die Analyse der „Slopes“ noch die Zeit-Ereignis-Analysen erforderten eine Imputation für fehlende Werte.</p> <p><u>Datenschnitt:</u> Zur Analyse des Ausmaßes des Zusatznutzens wird hauptsächlich die primäre Analyse nach 1:1 Paarbildung der Patienten der Studie 190-201/202 (Datenschnitt 15.10.2015) und 190-901 (historische Kontrolle; Datenschnitt Februar 2015) herangezogen, da nur diese eine Vergleichbarkeit der Gruppen hinsichtlich prognostisch relevanter Faktoren gewährleistet. Die Ergebnisse patientenrelevanter Endpunkte des finalen Studienberichts der Studie 190-201 (ohne „matching“) werden unterstützend dargestellt.</p>
190-202	<p>Bei der Studie 190-202 handelt es sich um eine Folgestudie der Studie 190-201 und die Operationalisierung des Endpunkts entsprach derjenigen der Studie 190-201. Analog der Studie 190-201 wurde der Endpunkt alle 8 Wochen (<math>\pm 3</math> Tage) während der Behandlung erhoben. Alle 24 Wochen (<math>\pm 3</math> Tage) wurde die Erhebung aufgezeichnet.</p>
190-901 Natürlicher Verlauf (Supplemental Report)	<p>Die Hamburg CLN2-Bewertungsskala (Summenscore bestehend aus der motorischen und Sprachfunktion sowie dem Sehvermögen; Domäne Krampfanfälle separat) wurde erstmals im Jahr 2002 in der Publikation von Steinfeld et al. beschrieben [5], allerdings wurde die Skala schon lange Zeit vorher in Hamburg eingesetzt wie die Studie 190-901 zum natürlichen Verlauf der Erkrankung zeigt. Die WCMC-Skala, die von einer unabhängigen Arbeitsgruppe entwickelt wurde, ist der HML-Skala hinsichtlich der Domänen „motorische Funktion“ und „Sprache“ vergleichbar. Die Graduierungen, um den Verlust essentieller motorischer sowie der Sprachfunktion quantitativ zu erfassen, sind sehr ähnlich (Tabelle 4-29). Die Hamburg-Skala (ohne Sehvermögen) wurde in einer Publikation aus dem Jahr 2008 zur Sicherheit und Wirksamkeit einer Gentherapie bei Patienten mit CLN2 eingesetzt [35].</p> <p>Der HML-Score war der primäre Endpunkt der Studie 190-901 (Supplemental Report) [56]. Die Operationalisierung des Endpunkts ist in Tabelle 4-29 dargestellt.</p> <p>Aufgrund des Designs der Studie (Patientenregister) wurde der HML-Score nicht standardisiert erhoben und es waren keine Erhebungszeitpunkte festgelegt. Die Daten wurden teilweise retrospektiv bzw. die aktuelleren Daten prospektiv erhoben.</p> <p>Die Imputationsmethoden für fehlende Werte sind oben beschrieben (Studie 190-201).</p> <p>Datenschnitt war Februar 2015 (Zeitpunkt des Datentransfers von DEM-CHILD).</p>

In Tabelle 4-40 sind mögliche Verzerrungsquellen des Endpunkts angegeben.

Tabelle 4-40: Bewertung des Verzerrungspotenzials des Endpunkts „Krankheitsprogression auf CLN2 spezifischen Bewertungsskalen (ML/HML-Skala)“ mit dem zu bewertenden Arzneimittel – weitere Untersuchungen

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
Cerliponase alfa-201/202	Nicht zutreffend	nein	ja	ja	ja	Nicht zutreffend

Das Verzerrungspotential auf Endpunktebene ist nur bei randomisierten Studien abschließend zu bewerten. Dennoch können gemäß IQWiG (Methoden 4.2) auch „Studien trotz niedriger Qualität im Einzelfall für bestimmte Endpunkte Resultate hoher Ergebnissicherheit liefern“ [23].

Bei nicht randomisierten Studien ist insbesondere die Vergleichbarkeit der Interventionsgruppen bzw. adäquate Berücksichtigung prognostisch relevanter Faktoren für die Bewertung des Verzerrungspotentials von Bedeutung (IQWiG Allgemeine Methoden 4.2) [23]. Zur Paarbildung zwischen den Teilnehmern der Studie 190-201/202 und der historischen Kontrolle aus unbehandelten Patienten wurden verschiedene Algorithmen eingesetzt, um hinsichtlich bekannter Prognosefaktoren eine möglichst hohe Vergleichbarkeit der Gruppen zu erreichen. Zudem wurden verschiedene Sensitivitätsanalysen durchgeführt, um die Robustheit der Ergebnisse zu belegen. Dennoch kann, wie auch bei jeder anderen, nicht kontrollierten Studie, ein Bias durch unbekannte Störfaktoren nicht ausgeschlossen werden. Eine Ausnahme bildet der sogenannte „dramatische“ Effekt einer Intervention (z. B. RR von 5-10 auf einem Signifikanzniveau von  $\leq 1\%$ ), der durch Störgrößeneinflüsse allein nicht mehr plausibel erklärt werden kann [23].

Der Beleg eines „dramatischen“ Effekts kann - selbst auf Basis einer nicht kontrollierten, klinischen Studie - zur Bewertung des Zusatznutzens ausreichend sein, vorausgesetzt, der schicksalhafte Verlauf der Erkrankung ist gut dokumentiert [23]. Bei einigen Patienten der historischen Kontrollgruppe (Studie 190-901) enthielten die Datensätze sowohl prospektiv als auch retrospektiv erhobene Daten (basierend auf Videoanalysen, etc.). Dies kann zu Verzerrungen führen, allerdings ist zu berücksichtigen, dass es sich bei der CLN2 um eine seltene Erkrankung handelt und die DEM-CHILD-Datenbank die derzeit höchste Evidenz mit Längsschnittdaten darstellt. Die Tatsache, dass der Krankheitsverlauf in anderen, unabhängigen Studien sehr ähnlich war (z. B. die Querschnittsstudien des WCMC), stellt sicher, dass DEM-CHILD den natürlichen Krankheitsverlauf widerspiegelt und daher als historische Kontrolle valide ist.

Das ITT-Prinzip wurde in der Studie 190-201 angemessen umgesetzt. Der Ausschluss eines Patienten, der seine Einwilligung nach nur einer Infusion aus Gründen zurückzog, die nicht mit



der Wirksamkeit oder Sicherheit in Zusammenhang standen, war wissenschaftlich begründet. Alle Patienten der ITT-Population schlossen die Nachbeobachtung vollständig ab und wurden in die Studie 190-202 aufgenommen.

Alle Prüferärzte wurden durch einen erfahrenen Experten darin ausgebildet, wie die CLN2-Skalen aufgebaut und anzuwenden sind. Die Erhebungen wurden in den Studienzentren regelmäßig überwacht, um zu gewährleisten, dass diese zwischen und innerhalb der Zentren einheitlich waren. Nach einem Handbuch der Cochrane-Gruppe und der AWMF (2016) ist eine Verzerrung durch fehlende Verblindung bei objektiven Parametern unwahrscheinlich [64]. Subjektive Aspekte spielen bei den Erhebungen der ML-Skalen wahrscheinlich zumindest eine geringe Rolle, allerdings stellt jede Stufe auf der ML-Skala eine leicht erkennbare, wesentliche Änderung der Funktion dar. Die Bewertungen der ML-Scores (0-6) durch einen unabhängigen Experten stimmten nahezu perfekt mit denjenigen der Prüferärzte der Studie 190-201 überein, sodass hinsichtlich der ML-Scores von einer hohen Interrater-Reliabilität ausgegangen werden kann. Daher wird nicht angenommen, dass subjektive Interpretationen seitens der einzelnen Beurteiler die Erhebung der ML-Scores entscheidend beeinflussen.

Um die beiden Kohorten vergleichen zu können, ist Voraussetzung, dass die HML-Skala (historische Kontrolle, Studie 190-901) und die in der Studie 190-201/202 eingesetzte, modifizierte ML-Skala einander sehr ähnlich sind. Die ML-Skala wurde, verglichen mit der HML-Skala (historische Kontrolle), nur geringfügig verändert. Ziel dieser Änderungen waren objektive Ankerpunkte, die gewährleisten sollten, dass die Bewertungen zwischen und innerhalb der Zentren konsistent sind. Da es sich bei der Studie 190-901 um eine historische Kontrolle handelt und eine Videoüberwachung in dieser Studie nicht Teil der klinischen Praxis war, konnten die Daten der historischen Kontrolle nicht nachträglich mittels der ML-Skala (Cerliponase alfa-Studien) analysiert werden. Daher wurde eine Brückenstudie durchgeführt, um auf Grundlage der Videoaufzeichnungen der Studien 190-201 und 190-202 zu untersuchen, ob die HML- (bei der historischen Kontrolle eingesetzt) und ML-Skala vergleichbar sind. Zu allen Untersuchungszeitpunkten (Baseline, Woche 25, 48 und 72) war die Übereinstimmung zwischen den beiden Bewertungen (motorisch und sprachlich kombiniert) beträchtlich ( $0.67 \leq K\omega \leq 0.79$ ) [24]. Die „Meilensteine“ der Entwicklung der Kinder, die beide Skalen erfassen, blieben dabei unverändert. Daher erscheint es unwahrscheinlich, dass die geringfügigen Änderungen der ML- im Vergleich zur HML-Skala zu einer relevanten Verzerrung führen könnten.

Während der stabilen Dosisphase wurde die ML-Skala alle 8 Wochen über 48 Wochen erhoben. In Anbetracht des durchschnittlichen Abfalls des HML-Scores um etwa zwei Punkten pro Jahr bei unbehandelten Patienten, sind die Erhebungsintervalle sowie der Zeitraum ausreichend, um die Progression sowie Rate der Progression bei Patienten mit milder bis mittelschwerer CLN2 valide erfassen zu können. Aufgrund des Designs waren in der Studie 190-901 (natürlicher Verlauf) Erhebungszeitpunkte oder Visiten nicht prospektiv festgelegt. BioMarin hat Anstrengungen unternommen, um möglichst alle Daten, die nicht auf direkten klinischen Beobachtungen beruhen (imputierte Werte), aus der Datenbank zu löschen, und methodisch sinnvolle Verfahren eingesetzt, um fehlende Werte zu ersetzen.

Bei dem primären Wirksamkeitsendpunkt der Studien 190-201/202 handelt es sich um einen kombinierten Endpunkt, der sowohl die motorischen als auch sprachlichen Fähigkeiten (ML) erfasst. Gemäß IQWiG „Allgemeine Methoden“ sind kombinierte Endpunkte mit Vorsicht zu interpretieren. Das IQWiG nimmt hierzu wie folgt Stellung: „Eine Voraussetzung für die Berücksichtigung eines kombinierten Endpunkts ist, dass die einzelnen Komponenten des kombinierten Endpunkts sämtliche im Berichtsplan definierte patientenrelevante Endpunkte darstellen. Die Komponenten sollen von ähnlicher „Schwere“ sein. D. h. nicht, dass sie von identischer Bedeutung sein müssen“ [23]. Die Fähigkeit, sich selbständig fortzubewegen, und der Erwerb der Sprache bilden jeweils „Meilensteine“ in der Entwicklung des Zentralnervensystems ab und sind für das menschliche Leben von grundlegender Bedeutung; daher haben beide Aspekte für die Patienten und deren Eltern/Betreuer eine ähnliche, allerdings wahrscheinlich nicht identische Bedeutung. Bei beiden Domänen handelt es sich um weitgehend objektive Parameter (basierend auf Beobachtungen) und diese wurden standardisiert erhoben. Das Studienprotokoll enthält plausible Erklärungen, warum die motorische und Sprachdomäne kombiniert wurden und andere Domänen, die für die Kinder oder deren Familien wahrscheinlich ebenfalls relevant sind (wie z. B. Sehvermögen und Krampfneigung), nicht in den primären Wirksamkeitsendpunkt einbezogen wurden. Sowohl die Einzelkomponenten als auch der Gesamtscore wurden prospektiv ausgewertet, wie seitens des IQWiG empfohlen [23]. Zusätzlich wurden die 9-Punkt-MLV-Skala (plus Sehvermögen) und 12-Punkt-MLVS-Skala (plus Sehvermögen und Krampfanfälle) analysiert.

Die Teilnahme an einer klinischen Studie und die damit verbundenen häufigen Untersuchungen könnten dazu geführt haben, dass bei den Teilnehmern der Studie, verglichen mit der historische Kontrolle, die anti-epileptische sowie andere Therapien optimiert wurden. Dies könnte die Ergebnisse des CLN2-Gesamt-Scores verzerren und war ein Grund, warum die Domänen „Krampfanfälle“ und „Myoklonus“ nicht in den primären Endpunkt einbezogen wurden. Begleitmedikationen haben wahrscheinlich keinen Einfluss auf den ML-Score, da die üblicherweise eingesetzten Therapien rein symptomatisch wirken und den Krankheitsverlauf nicht verändern. Daher wird nicht angenommen, dass Begleitmedikationen die motorischen oder sprachlichen Fähigkeiten nennenswert oder nachhaltig verändern.

Im Studienverlauf wurde die Primäranalyse der Krankheitsprogression, gemessen mittels ML-Skala, von einer Slope- in eine Responder-Analyse geändert (Version 3 des SAPs, 25.03.2016). Responder waren als Patienten definiert, deren ML-Score innerhalb von 48 Wochen nicht dauerhaft um  $\geq 2$  Punkte abfiel oder die in Woche 48 nicht einen ML-Wert = 0 hatten (Studientag 340 bezogen auf die erste 300 mg-Infusion). Mit dieser Änderung wurden Empfehlungen der Zulassungsbehörden entsprochen und der SAP lag bereits vor Öffnen der Datenbank in der finalen Version vor.

Darüber hinaus wurde im Rahmen der ISE ein historischer Vergleich („matched“) zur Analyse des primären Endpunkts nach Studienbeginn eingeführt (ISE SAP). Bereits in Amendment 2 des Studienprotokolls (datiert 07.03.2013; gültig bei Einschluss des ersten Patienten) war ein historischer Vergleich auf Grundlage existierender CLN2-Register geplant. Responder waren als Patienten definiert, deren ML/HML-Score sich innerhalb von 48 Wochen um weniger als

zwei Punkte verschlechterte. Mit diesen Änderungen wurden ebenfalls Empfehlungen der Zulassungsbehörden umgesetzt. Die ursprünglich geplante Slope-Analyse blieb als wichtiger Endpunkt erhalten und wurde wie vorgesehen analysiert. Daher liefern diese Änderungen keine Hinweise auf eine ergebnisgesteuerte Berichterstattung.

Hinsichtlich der Responder-Definition zur primären Analyse (kein Abfall um zwei Punkte und ML-Score  $\neq 0$  in Woche 48 im SAP und Abfall um weniger als zwei Punkte / 48 Wochen im ISE SAP; 1:1 Paarbildung mit der historischen Kontrolle) ist zu berücksichtigen, dass bereits die Fallzahlschätzung auf einem Abfall um zwei Punkte innerhalb von 18 Monaten nach Auftreten von Symptomen beruhte, basierend auf Längsschnittdaten bei 30 Patienten mit genetisch gesicherter CLN2 (Amendment 2 des Studienprotokolls vom 07.03.2013; gültig bei Einschluss des ersten Patienten).

Zusammenfassend ist festzuhalten, dass für den primären Endpunkt der Studie 190-201/202 ein historischer Vergleich durchgeführt wurde, bei dem die relevanten Prognosefaktoren zur Paarbildung herangezogen wurden, um die Vergleichbarkeit der Gruppen zu gewährleisten. Es wurden verschiedene andere Algorithmen zur Paarbildung eingesetzt, um den primären Algorithmus zu bestätigen. Gemäß IQWiG kann der Nutzen einer medizinischen Intervention auch durch die Beobachtung einer Umkehr des (quasi-)deterministischen Verlaufs bei einer gut dokumentierten Serie von Patientinnen und Patienten belegt werden, vorausgesetzt, der Verlauf einer Erkrankung ist sicher oder nahezu sicher vorhersagbar und es existieren keine krankheitsmodifizierenden Behandlungsoptionen. Voraussetzung ist der Nachweis eines „dramatischen“ Effekts, der nicht mehr allein durch Störgrößen erklärbar ist [23].

Die Ergebnisse der Studien 190-201/202 werden in Übereinstimmung mit dem TREND-Statement berichtet [19], welches international den Standard für nicht randomisierte Studien darstellt und daher für die Studien 190-201/202 maßgeblich ist. Nach Item 17 des TREND-Statements sollten die Ergebnisse aller primären und sekundären Endpunkte dargestellt werden. Diese Vorgabe wird in dem vorliegenden Dossier nicht vollständig umgesetzt, da nur patientenrelevante Endpunkte berücksichtigt werden.

#### **4.4.2.3.1.1 Mittlere Änderung der ML/HML-Skala von Baseline in den Studien 190-201/202 und 190-901 (historische Kontrolle); deskriptive Analyse**

Tabelle 4-41 zeigt die Baseline-Werte sowie deskriptive Analysen der mittleren Änderung der ML/HML-Werte im Studienverlauf bei den behandelten Patienten der Studie 190-201/202 und unbehandelten Patienten der Studie 190-901 (historische Kontrolle).

Alle Daten beziehen sich auf die ITT-Population und das 300 mg-Dosisintervall der Studie 190-201/202. Baseline entspricht hierbei der letzten Beobachtung vor der ersten Infusion von 300 mg.

Für die Studie 190-901 sind die Ergebnisse sowohl der „auswertbaren“ Population ohne Paarbildung (n=41) als auch nach Paarbildung mit den Patienten der Studie 190-201 dargestellt. Die mittleren ML/HML-Werte waren bei den Patienten der Studie 190-901 ohne Paarbildung

etwas höher (besser) als bei den behandelten Patienten. Allerdings nahm der ML/HML-Score in allen Analysen, unabhängig von der eingesetzten Methode, bei den unbehandelten Patienten der historischen Kontrolle im Mittel um etwa zwei Punkte ab verglichen mit 0,4-0,5 Punkten bei den behandelten Patienten.

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

Tabelle 4-41: Geschätzte Abnahme des ML/HML Wertes über 48 Wochen: Cerliponase alfa (Studie 190-201) vs. unbehandelte Kontrolle (Studie 190-901), mit oder ohne Paarbildung

Baseline Schätzung 190-901 Analysepopulation 190-201	Studie (n)	Mittelwert (± SD) der Punkte auf der ML/HML-Skala			
		Baseline-Wert	Wert in Woche 25	Wert in Woche 49	Änderung von Baseline bis Woche 49
<i>Ohne Paarbildung</i>					
Schätzung durch Interpolation, Mittlerer Wert <sup>1</sup>	190-901 (n=41)	4,4 (0,78) <sup>1</sup>	3,4 (0,94)	2,4 (1,17) (n=38)	-2,1 (1,09) (n=38)
ITT	190-201 (n=23)	3,5 (1,20) <sup>1</sup>	3,1 (1,34)	3,0 (1,33)	-0,4 (0,79)
Schätzung durch Interpolation <sup>2</sup>	190-901 (n=41)	4,4 (0,78)	3,6 (1,01)	2,7 (1,19) (n=38)	-1,8 (1,10) (n=38)
ITT	190-201 (n=23)	3,5 (1,20)	3,1 (1,34)	3,0 (1,33)	-0,4 (0,79)
Schätzung unter Einsatz eines Zeitfensters („windowing“), Mittlerer Wert <sup>1</sup>	190-901 (n=41)	4,4 (0,78)	3,4 (0,93) (n=17)	2,3 (1,39) (n=27)	-2,3 (1,30) (n=27)
ITT	190-201 (n=23)	3,5 (1,20)	3,1 (1,36)	3,0 (1,33)	-0,4 (0,79)
Schätzung durch Interpolation („Wirksamkeitspopulation“) ML <6, Mittlerer Wert	190-901 (n=41)	4,4 (0,78)	3,4 (0,94)	2,4 (1,17) (n=38)	-2,1 (1,09) (n=38)
ITT <sup>1,6</sup>	190-201 (n=21)	3,2 (0,94)	2,8 (1,04)	2,8 (1,00)	-0,5 (0,82)
<i>Mit Paarbildung</i>					
Schätzung durch Interpolation, „many-to-one“ Paarbildung (ML/HML Score) <sup>3</sup> , Mittlerer Wert	190-901 (n=23)	3,5 (1,19)	2,2 (1,16)	1,5 (1,05)	-2,0 (0,29)
ITT <sup>1</sup>	190-201 (n=23)	3,5 (1,20)	3,1 (1,34)	3,0 (1,33)	-0,4 (0,79)

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

Baseline Schätzung 190-901 Analysepopulation 190-201	Studie (n)	Mittelwert (± SD) der Punkte auf der ML/HML-Skala			
		Baseline-Wert	Wert in Woche 25	Wert in Woche 49	Änderung von Baseline bis Woche 49
Schätzung durch Interpolation, „many-to-one“ Paarbildung (ML/HML Score und Alter) <sup>4</sup> , Mittlerer Wert ITT <sup>1,4</sup>	190-901 (n=22)	3,4 (1,09)	2,1 (1,09)	1,4 (1,06)	-2,0 (0,69)
	190-201 (n=22)	3,4 (1,09)	3,0 (1,21)	2,9 (1,19)	-0,4 (0,80)
Schätzung durch Interpolation, „many-to-one“ Paarbildung (ML/HML Score und Genotyp) <sup>5</sup> ITT <sup>1</sup>	190-901 (n=23)	3,5 (1,19)	2,3 (1,39)	1,7 (1,37)	-1,8 (0,47)
	190-201 (n=23)	3,5 (1,20)	3,1 (1,34)	3,0 (1,33)	-0,4 (0,79)

<sup>1</sup> 190-201: Baseline ist der letzte Wert vor der ersten 300 mg-Infusion; 190-901: Mittlerer Wert einer Reihe aufeinander folgender Werte (≤5) nach dem ersten Wert <6.  
<sup>2</sup> 190-201: Baseline ist der letzte Wert vor der ersten 300 mg-Infusion; 190-901: Baseline ist der erste Wert <6.  
<sup>3</sup> Paarbildung auf Basis des HML/ML-Scores zu Baseline.  
<sup>4</sup> Paarbildung auf Basis des HML/ML-Scores und Alters zu Baseline; ein Patient aus 190-201 konnte nicht gepaart werden.  
<sup>5</sup> Paarbildung auf Basis des HML/ML-Scores zu Baseline und des Genotyps (häufige versus seltene Allele).  
<sup>6</sup> Ausschluss von zwei Patienten (190-201/202) mit stabilem ML-Score von 6.  
 Positive Werte entsprechen einer Verbesserung und negative Werte einer Verschlechterung des klinischen Zustands.  
 ISE: Tabelle 2.1.1, Tabelle 2.1.1.a, Tabelle 2.1.2, Tabelle 2.1.3, Tabelle 2.1.5, Tabelle 2.1.6, Tabelle 2.1.7

Abbildung 11 zeigt die Änderung auf der 0-6-Punkt ML/HML-Skala im Zeitverlauf bei den auswertbaren Patienten der Studie 190-901 sowie Patienten der Studie 190-201. Bei den Patienten der historischen Kontrolle nahmen die mittleren Scores auch über 48 Wochen hinaus linear ab, während der Score bei den behandelten Patienten im Studienverlauf nahezu unverändert stabil blieb.

Wie die deskriptiven Analysen der 0-9-Punkt MLV/HMLV-Skala (zusätzliche Domäne Sehvermögen; Abbildung 12) und 0-12-Punkt MLVS/HMLVS-Skala (zusätzliche Domänen Sehvermögen und Krampfanfälle; Abbildung 13) zeigen, ändern sich die mittleren Scores über die 48-wöchige Behandlung in ähnlicher Weise. Diese Ergebnisse zeigen, dass die motorische und Sprachdomäne die Krankheitsprogression unter den gegebenen Studienbedingungen am sensitivsten erfassen. Der Einschluss der zusätzlichen Domänen zum Sehvermögen und der Krampfaktivität trägt während der vorgegebenen Behandlungsdauer von 48 Wochen nicht wesentlich zur Messung des Behandlungseffekts bei. Dennoch zeigt sich bei allen klinischen Domänen, die durch die CLN2-spezifischen Bewertungsskalen erfasst werden, ein dauerhafter Behandlungseffekt: Motorische sowie sprachliche Fähigkeiten, Sehvermögen und Krampfanfälle (Abbildung 13).

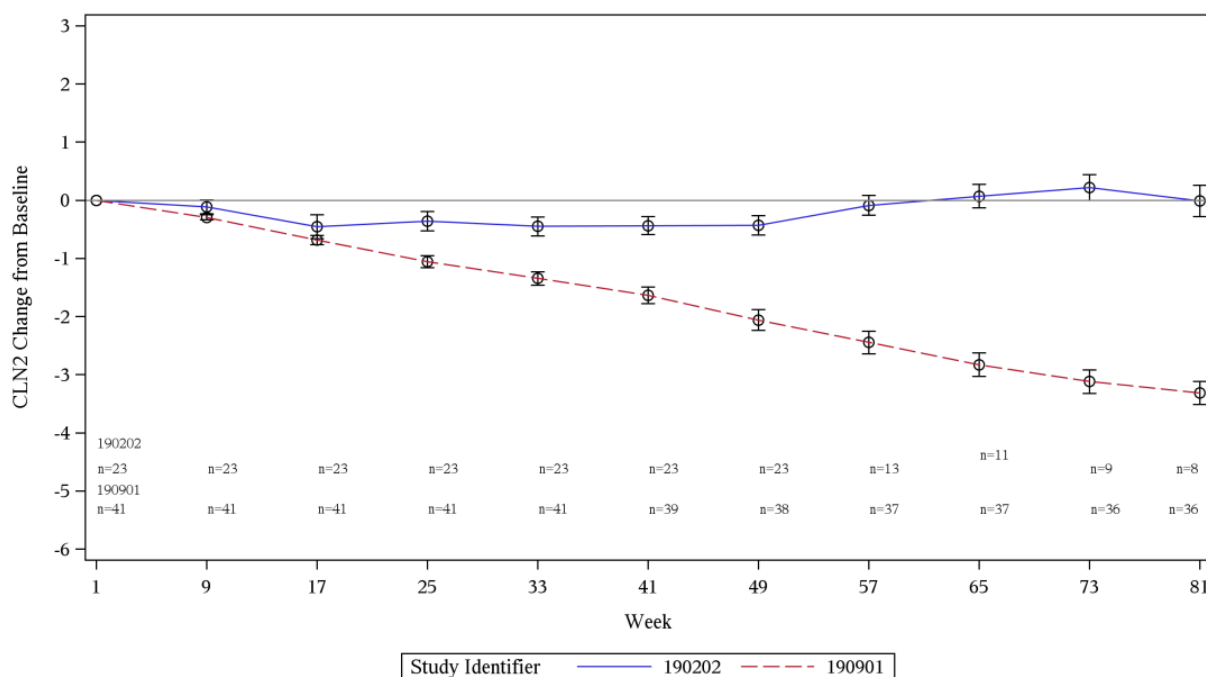


Abbildung 11: Mittlere Änderung der ML/HML-Skala bezogen auf Baseline (Studie 190-201/202, ITT-Population; 300 mg-Dosisintervall / Studie 190-901 auswertbare Patienten)

Die senkrechten Balken bilden den Standardfehler des Mittelwerts ab. Beide Baseline-Werte wurden auf 0 gesetzt, um die relative Änderung, bezogen auf Baseline, vergleichen zu können. Die ML/HML-Scores wurden bei jedem Patienten der Studie 190-201/202 sowie jedem Patienten der Studie 190-901 zu denselben Zeitpunkten eingegeben. Dies ermöglichte eine deskriptive Analyse der ML/HML-Scores zu gemeinsamen Zeitpunkten, bei der jeder Patient zur deskriptiven Statistik beitrug.

Quelle: ISE Abbildung 2.1.10

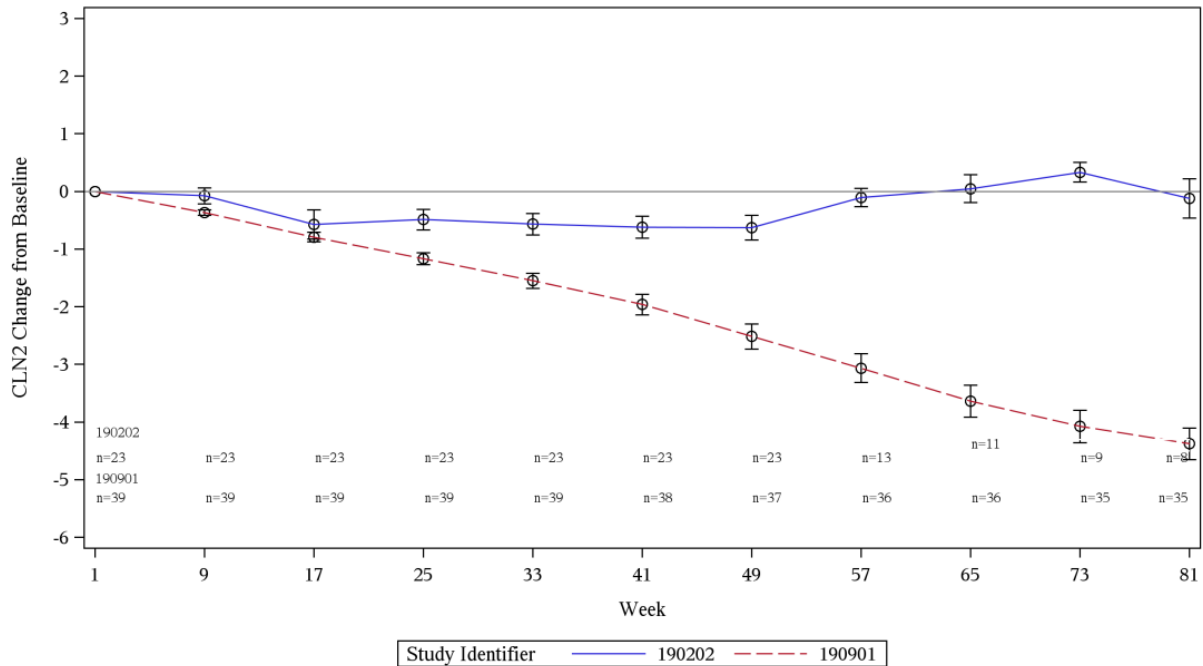


Abbildung 12: Mittlere Änderung der MLV/HMLV-Skala bezogen auf Baseline (Studie 190-201/202, ITT-Population; 300 mg-Dosisintervall / Studie 190-901 auswertbare Patienten)

Quelle: ISE Abbildung 2.2.6

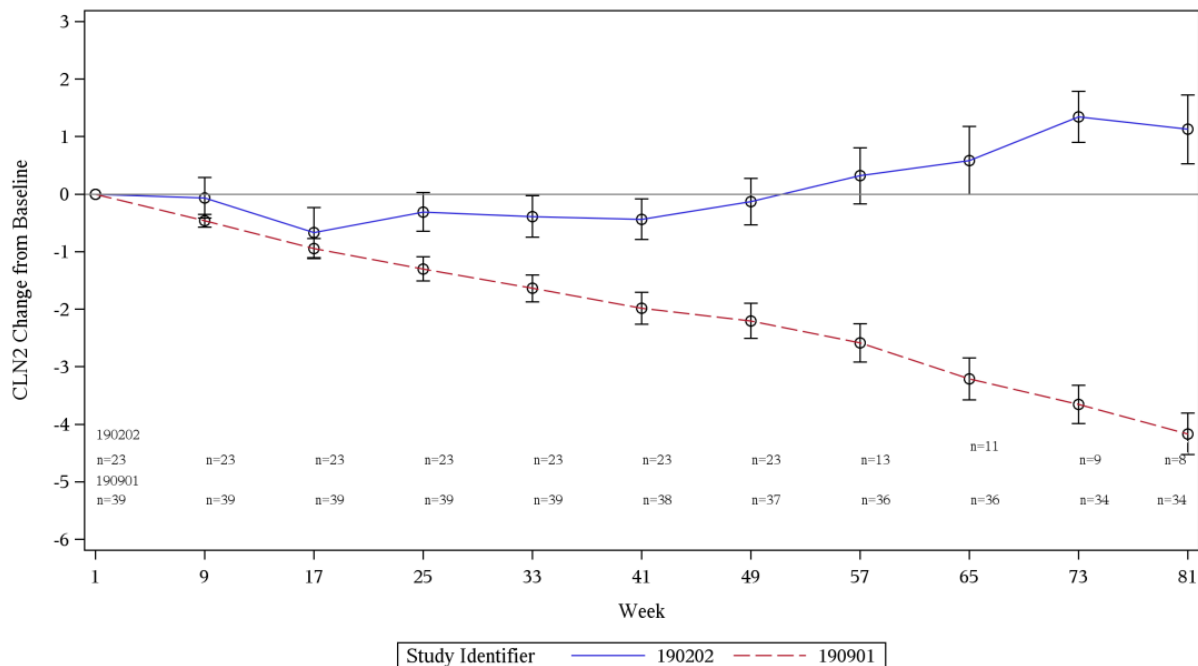


Abbildung 13: Mittlere Änderung der Gesamt-Scores der MLVS/HMLVS-Skala bezogen auf Baseline (Studie 190-201/202, ITT-Population; 300 mg-Dosisintervall / Studie 190-901 auswertbare Patienten)

Quelle: ISE Abbildung 2.3.6



#### 4.4.2.3.1.2 Patientenrelevante Änderung der ML/HML-Bewertungsskala

Bei den Responder-Analysen des primären Endpunkts war ein Ansprechen auf die Therapie wie folgt definiert: Änderung um  $<2$  Punkte auf der ML/HML-Bewertungsskala (motorische und Sprachdomäne kombiniert). Kein Therapieansprechen war definiert als eine Verschlechterung um  $\geq 2$  Punkte. Seitens der EMA wurde auf Basis des natürlichen Verlaufs der Erkrankung eine jährliche Abnahme um 2 Punkte als patientenrelevantes Kriterium abgelehnt.

Eine Verschlechterung um mehr als zwei Punkte auf der krankheitsspezifischen Bewertungsskala ist zweifelsohne für Patienten mit CLN2 relevant (siehe Tabelle 4-29). Um die Patientenrelevanz des gewählten Erfolgskriteriums formal zu unterstützen und um abschätzen zu können, welcher Unterschied für Patienten mit CLN2 relevant ist (MID, „minimal important difference“), wurde ein verteilungsbasiertes, statistisches Verfahren eingesetzt. Das 0,5-Fache der Standardabweichung (SD) des Baseline-Wertes wurde als patientenrelevante Änderung vorgeschlagen [65].

Dividiert man die SD des Baseline-Wertes aus der Studie 190-201 (siehe Tabelle 4-41) durch zwei (d. h. 1,2 geteilt durch 2) ergibt dies 0,6 Punkte. Daher kann ein Wert von 0,6 Punkten auf der eingesetzten Krankheitsskala als MID interpretiert werden.

Das eingesetzte Messinstrument erfasst klinisch klar unterscheidbare „Meilensteine“ und dies stützt zusammen mit der oben angegebenen MID nachdrücklich den Einsatz des Schwellenwerts von 2 Punkten, um in der primären Wirksamkeitsanalyse patientenrelevante Effekte abzubilden. Der zur Primäranalyse eingesetzte Schwellenwert ( $<2/\geq 2$  Punkte) ist 3,33-fach höher als die mittels eines verteilungsbasierten, statistischen Verfahrens berechnete MID.

Bislang unveröffentlichte Ergebnisse aus der historischen Studie zeigten, dass der Gesamtscore der Hamburg Skala (vier Dimensionen) oder der Weill-Cornell-Skala (vier Dimensionen) signifikant mit dem Gesamtscore des PedsQL™ (Family Impact Modul) korrelieren.

#### 4.4.2.3.1.3 Primäre Analyse: Abfall des CLN2-Scores um zwei Punkte (Responder- und Ereigniszeitanalysen)

##### Responder-Analyse (1:1 Paarbildung mit der historischen Kontrolle)

Zusammen mit der Ereigniszeitanalyse stellte die Responder-Analyse den primären Endpunkt der Studie dar. Patienten galten als Therapieresponder, wenn der ML/HML-Score innerhalb von 48 Wochen um weniger als zwei Punkte abfiel. Analysiert wurden die Patienten der ITT-Population der Studie 190-201/202, denen Patienten der historischen Kontrolle im Verhältnis 1:1 zugeteilt wurden (1:1 „matching“), vorausgesetzt, der ML/HML-Wert zu Baseline stimmte exakt überein und der Altersunterschied war so gering wie möglich (allerdings nicht größer als 12 Monate) [58].

Die geschätzte Ansprechrate (Abfall des ML/HML-Scores um  $<2$  Punkte) betrug 91% ( $n=2/20$ ) bei den Patienten der Studie 190-201/202 (ITT) verglichen mit 45% ( $n=10/22$ ) bei den gepaarten Patienten der historischen Kontrolle. Die Ansprechrate unterschied sich im

Gruppenvergleich um 46% zugunsten von Cerliponase alfa ( $p=0,0028$ ) und daher konnte die Nullhypothese verworfen werden (Tabelle 4-42). Das relative Risiko (RR) betrug hinsichtlich der Responder-Rate  $RR=2,00$  ([95% KI: 1,2420; 3,2207];  $p=0,0435$ ) und hinsichtlich der Non-Responder-Rate  $RR=0,167$  ([95% KI: 0,04212; 0,65942];  $p=0,01067$ ), jeweils zugunsten der Behandlung mit Cerliponase alfa. Dieses Ergebnis zeigt, dass die Erkrankung bei deutlich mehr Patienten der historischen Kontrolle („matched“) als der Cerliponase alfa-Gruppe voranschreitet (ausgedrückt als Non-Responder auf der ML/HML-Skala; siehe Tabelle 4-42).

Tabelle 4-42: Primäre Analyse der Ansprechrate gemessen mittels ML/HML-Skala ( $<2$  oder  $\geq 2$  Punkte/48 Wochen): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) nach 1:1 Paarbildung [57]

Response	Studie		Differenz (190-201) - (190-901)	p-Wert <sup>1</sup> Differenz	190-201/190-901		
	190-901 (n=22)	190- 201/202 (n=22)			RR	95% KI	p-Wert
Ja (Rate $<2$ )	10 (45%)	20 (91%)	46%	0,0028	2,000	1,2420; 3,2207	0,0044
Nein (Rate $\geq 2$ )	12 (55%)	2 (9%)	-	-	0,16667	0,04212; 0,65942	0,01067

<sup>1</sup> Exakter Test nach Fisher

KI: Konfidenz Intervall; RR: relatives Risiko

Kriterien zur 1:1 Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq 12$  Monate zu Baseline (300 mg-Baseline in Studie 190-201). Bei Patienten der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1 Paarbildung.

Geschätzte Neigung (Slope; Intra-Subjekt) kategorisiert nach  $<2$  und  $\geq 2$  Punkte auf der ML/HML-Skala

ISE: Tabelle 3.1.1

Modul 2.7.3: Tabelle 2.7.3.3.2.2.1

### ***Sensitivitätsanalysen (Responder-Rate) - 1:1 Paarbildung mit der historischen Kontrolle***

Wie vorab festgelegt wurden 1.000 Simulationen zur Paarbildung eingesetzt, bei denen, verglichen mit der Primäranalyse, weniger komplexe Methoden zur Paarbildung verwendet wurden. Bei jeder Simulation wurden die Patienten der Studie 190-201/202 nach dem Zufallsprinzip (randomisiert) angeordnet und anschließend nacheinander, im Verhältnis 1:1 den Patienten der Studie 190-901 zugeteilt (Kriterien zur Paarbildung: Übereinstimmender ML/HML-Score, kleinster Altersunterschied und Altersunterschied  $\leq 12$  Monate). Die Responder-Analyse wurde für jede der 1.000 Simulationen wiederholt (siehe Tabelle 4-43). Die über 1.000 Simulationen ermittelte, mittlere Ansprechrate (Responder-Rate; SD) betrug 49,1% (4,85) bei den Patienten der historischen Kontrolle (auswertbare Patienten) verglichen mit 90,8% (0,21) bei den Patienten der Cerliponase alfa-Gruppe (ITT, 300 mg-Dosisintervall). Daraus ergibt sich ein mittlerer Unterschied von -41,7% (4,83) bei p-Werten zwischen 0,001 und 0,033 (einseitig, exakter Test nach Fisher; Tabelle 4-43). Über alle 1.000 Simulationen hinweg betrug die Ansprechrate bei den behandelten Patienten mindestens 90,0% verglichen mit höchstens 61,9% bei den unbehandelten Patienten (ebenfalls 1.000 Simulationen). Die

Ergebnisse dieser Responder-Analysen, denen 1.000 Simulationen zugrunde lagen, waren der Primäranalyse vergleichbar.

Tabelle 4-43: Ansprechrate gemessen mittels ML/HML-Skala (<2 oder  $\geq 2$  Punkte/48 Wochen) über 1.000 Simulationen zur 1:1 Paarbildung („matching“): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten; Sensitivitätsanalyse) [57]

Responder-Rate <sup>1</sup> (%)	Studie 190-901	Studie 190-201/202	Differenz (190-901) - (190-201/202)	Einseitiger p-Wert <sup>2</sup>
N	1.000	1.000	1.000	1.000
Mittelwert (SD)	49,1 (4,85)	90,8 (0,21)	-41,7 (4,83)	0,0039
Median	50,0	90,9	-40,9	0,003
25 <sup>th</sup> ; 75 <sup>th</sup> Perzentile	45,5; 52,4	90,5; 90,9	-45,5; -38,1	0,001; 0,007
Minimum, Maximum	33,3; 61,9	90,0; 90,9	-57,1; -28,6	<0,0001; 0,033

<sup>1</sup> Geschätzte Rate (Intra-Subjekt) des Abfalls auf der ML/HML-Skala von <2 Punkten über 48 Wochen.  
<sup>2</sup> Exakter Test nach Fisher  
Methode: Vorher-/Nachher-Vergleich (First Point/Last Point)  
Kriterien zur Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq 12$  Monate zu Baseline (300 mg-Baseline in Studie 190-201).  
Modul 2.7.3: Tabelle 2.7.3.3.2.2.2

Zur Überprüfung der Robustheit der Primäranalyse wurden verschiedene, weitere Sensitivitätsanalysen durchgeführt.

In den Tabellen werden folgende Abkürzungen verwendet, um die Art der Sensitivitäts (Subgruppen)-Analysen zu beschreiben:

- “Wirksamkeit”; Ausschluss von zwei Patienten, deren ML-Score im Verlauf der Studie unverändert bei sechs Punkten: “Ohne ML = 6” (n=20)
- Wie oben plus Ausschluss eines Paares mit einem 300 mg-Baseline ML-Score = 1: “Ohne ML = 1” (n=19)
- Erweiterung eines 1:1 Paarbildungskriteriums (Altersunterschied zu Baseline) auf 21 Monate: “21 Monate” (n=23)
- Änderung des Paarbildungs-Algorithmus, sodass die Quadratsumme des Abstands zwischen ML/HML-Score und Alter (in Monaten) zu Baseline  $\leq 36$  und das mittlere Abstandsquadrat möglichst niedrig ist (n=23); bei diesem Algorithmus konnten sich die ML/HML-Scores zu Baseline um bis zu einen Punkt in beide Richtungen unterscheiden (d. h. der Score konnte bei einem Patienten der Studie 190-201/202 um einen Punkt höher oder niedriger sein als bei dem entsprechenden Patienten der Studie 190-901)

Die *a priori* definierten Sensitivitäts-Analysen bestätigten die Ergebnisse der Primäranalyse und die Ansprechrate (Responder) unterschied sich zwischen den behandelten Patienten (190-201/202) und den Patienten der historischen Kontrolle (190-901) statistisch signifikant. Das relative Risiko der Responder-Raten (Abfall um  $<2$  Punkte/48 Wochen) lag im Bereich von zwei zugunsten von Cerliponase alfa und alle Sensitivitätsanalysen waren auf einem Niveau von  $p < 0,01$  signifikant. Bei allen Sensitivitätsanalysen lag das RR für Non-Response (Abfall um  $\geq 2$  Punkte/48 Wochen) unter 0,2 (Kehrwert  $> 5$ ), das Signifikanzniveau lag bei allen Vergleichen unter 0,02 (Tabelle 4-44).

Tabelle 4-44: Sensitivitätsanalysen der primären Responder-Analyse ( $< 2$  oder  $\geq 2$  Punkte / 48 Wochen) [57]

Population / Responder, Non-Responder	n	190-901 n (%)	190-201/202 n (%)	Differenz <sup>1</sup> (190-201)- (190-901)	p-Wert <sup>2</sup>	190-201/190-901		
						RR	95% KI	p-Wert
<i>Responder (&lt; 2 Punkte/48 Wochen gemessen mittels ML/HML-Skala)</i>								
ITT (Primär-analyse)	22	10 (45%)	20 (91%)	46%	0,0028	2,000	0,31050; 0,80517	0,00435
„Wirksamkeit“, ohne ML = 6 <sup>3</sup>	20	9 (45%)	18 (90%)	45%	0,0057	2,000	1,2057; 3,3175	0,0073
„Wirksamkeit“, ohne ML = 1 <sup>3,4</sup>	19	8 (42%)	17 (89%)	47%	0,0051	2,1250	1,2268; 3,6808	0,0072
ITT, 21 Monate <sup>5</sup> (ITT)	23	11 (48%)	21 (91%)	43%	0,0031	1,9091	1,2233; 2,9794	0,0044
Unterschiedlicher Algorithmus zur 1:1 Paarbildung, Methode #2 (ITT)	23	12 (52%)	21 (91%)	39%	0,0074	1,7500	1,1601; 2,6399	0,0076
<i>Non-Responder (<math>\geq 2</math> Punkte/48 Wochen gemessen mittels ML/HML-Skala)</i>								
ITT (Primär-analyse)	22	12 (55%)	2 (9%)	n.k.	n.k.	0,16667 iRR: 6,000	0,04212; 0,65942	0,01067
„Wirksamkeit“, ohne ML = 6 <sup>3</sup>	20	11 (55%)	2 (10%)	n.k.	n.k.	0,1818 iRR: 5,501	0,0461; 0,7179	0,0150
„Wirksamkeit“, ohne ML = 1 <sup>3,4</sup>	19	11 (58%)	2 (11%)	n.k.	n.k.	0,1818 iRR: 5,501	0,0464; 0,7216	0,0144
ITT, 21 Monate <sup>5</sup> (ITT)	23	12 (52%)	2 (9%)	n.k.	n.k.	0,1667 iRR: 5,999	0,0419; 0,6631	0,0110
Unterschiedlicher Algorithmus zur 1:1 Paarbildung, Methode #2 (ITT)	23	11 (48%)	2 (9%)	n.k.	n.k.	0,1818 iRR: 5,501	0,0452; 0,7310	0,0163
iRR: inverses relatives Risiko; n.k. nicht kalkuliert; RR relatives Risiko								

Population / Responder, Non- Responder	n	190- 901 n (%)	190- 201/202 n (%)	Differenz <sup>1</sup> (190-201)- (190-901)	p-Wert <sup>2</sup>	190-201/190-901		
						RR	95% KI	p-Wert
<sup>1,2</sup> Exakter Test nach Fisher <sup>3</sup> Ausschluss von n=2 mit stabilem ML-Score von 6 <sup>4</sup> Ausschluss eines Paares mit ML-Score von 1 zu Baseline <sup>5</sup> Erweiterung der 1:1 Paarbildungskriterien um Altersdifferenz von bis zu 21 Monaten. Wenn nicht anders angegeben wurden zur Paarbildung die Kriterien der Primäranalyse herangezogen: Identischer ML/HML-Score und Altersunterschied ≤12 Monate zu Baseline (300 mg-Baseline in Studie 190-201). Bei Patienten der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1 Paarbildung. Modul 2.7.3: Tabelle 2.7.3.3.2.2.3 ISE: Tabelle 3.1.2, Tabelle 3.1.3, Tabelle 3.1.6, Tabelle 3.1.8								

### *Einzeldomänen (motorische und Sprachdomäne getrennt)*

In ähnlichen Responder-Analysen nach 1:1 Paarbildung wurden die motorische und Sprachdomäne (bzw. Subskala) getrennt analysiert, wobei <1 und ≥1 Punkt als Response- bzw. Non-Response-Kriterium definiert waren. Der geschätzte Unterschied der Ansprechrate betrug auf der motorischen Subskala 41% (p=0,0097) und hinsichtlich der Sprachdomäne 66% (p<0,0001; Tabelle 4-45). Hervorzuheben ist, dass die Ansprechrate bei den Patienten der Cerliponase alfa-Gruppe bei beiden Einzeldomänen vergleichbar hoch war. Unter den unbehandelten Patienten der Kontrollgruppe, verschlechterte sich die Sprachdomäne bei etwas mehr Patienten als die motorische Domäne, wodurch sich die Ansprechrate hinsichtlich der Sprachdomäne, verglichen mit der motorischen Funktion, deutlicher unterschied. Die RR (95% KI) für Response bzw. Non-Response betrug auf der motorischen Subskala 1,900 (1,1676, 3,0919, p=0,0098) bzw. 0,2500 (Kehrwert der RR: 4; 0,0817, 0,7652, p=0,0151). Hinsichtlich der Sprachdomäne war die RR (95% KI) 3,800 (1,7461, 8,2696, p=0,0008) für Response und 0,1250 (Kehrwert der RR: 8; 0,0327, 0,4773, p=0,0023) für Non-Response.

Tabelle 4-45: Ansprechrate getrennt nach motorischer und Sprachdomäne auf der ML/HML-Skala (<1 oder  $\geq$ 1 Punkte/48 Wochen): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) nach 1:1 Paarbildung [57]

Domäne	n	190-901, n (%)		190-201/202, n (%)		Differenz p-Wert <sup>2</sup>	190-201/190-901		
		R	Non-R	R	Non-R		RR	95% KI	p-Wert
Motorische Domäne	22	10 (45%)	12 (55%)	19 (86%)	3 (14%)	41% 0,0097	R: 1,900	R: 1,1676; 3,0919	R: 0,0098
							Non-R: 0,2500	Non-R: 0,0817; 0,7652	Non-R: 0,0151
Sprachdomäne	21 <sup>1</sup>	5 (24%)	16 (76%)	19 (90%)	2 (10%)	66% <0,0001	R: 3,800	R: 1,7461; 8,2696	R: 0,0008
							Non-R: 0,1250	Non-R: 0,0327; 0,4773	Non-R: 0,0023

Non-R: Non-Responder; R: Responder; RR: relatives Risiko

<sup>1</sup> Ein Paar wurde ausgeschlossen, da der Baseline-Sprachscore bei mindestens einem der Patienten Null war.

<sup>2</sup> Exakter Test nach Fischer

Kriterien zur Paarbildung (matching): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq$ 12 Monate zu Baseline (300 mg-Baseline in Studie 190-201).

Geschätzte Neigung (Slope; Intra-Subjekt) kategorisiert nach <1 und  $\geq$ 1 Punkt auf der ML/HML-Skala (getrennt für die motorische und Sprachdomäne).

Modul 2.7.3: Tabelle 2.7.3.3.2.2.4

ISE: Tabelle 3.1.4, Tabelle 3.1.5

### Kaplan-Meier-Analyse der Krankheitsprogression:

#### Zeit bis zur dauerhaften Verschlechterung des ML/HML-Scores um $\geq$ 2 Punkte (1:1 Paarbildung mit der historischen Kontrolle)

Tabelle 4-46 zeigt die Ergebnisse der *a priori* definierten Kaplan-Meier-Analyse der Krankheitsprogression, ausgedrückt als Zeit bis der ML/HML-Score nachhaltig um  $\geq$ 2 Punkte oder auf den Wert von Null abfiel (d. h. der Wert fiel beispielsweise nicht um zwei Punkte ab, um anschließend wieder anzusteigen) nach 1:1 Paarbildung [58]. Zur Analyse wurde ein Cox-Modell eingesetzt, welches nach CLN2-Score und Alter zu Baseline adjustiert war.

Die Analyse zeigte einen sehr robusten Behandlungseffekt im Sinne einer deutlich verzögerten Krankheitsprogression unter Cerliponase alfa Behandlung (300 mg-Dosisintervall der Studie 190-201/202) im Vergleich zur historischen Kontrolle, gemessen als Zeit bis zur dauerhaften Verschlechterung des ML/HML-Scores um  $\geq$ 2 Punkte oder Verschlechterung auf den Wert von Null. Die HR betrug in der Primäranalyse 0,10 ([95% KI 0,03; 0,38],  $p=0,0005$ ) und stellt daher einen „dramatischen“ Effekt dar (Tabelle 4-46). Die unbehandelten Patienten (Studie 190-901)

hatten ein 10-fach höheres Risiko einer Krankheitsprogression (entsprechend der oben angegebenen Kriterien) als die mit Cerliponase alfa behandelten Patienten (Studie 190-201/202).

Tabelle 4-46: Kaplan-Meier-Analysen der Krankheitsprogression gemessen mittels ML/HML-Skala: Cerliponase alfa (190-201) vs. unbehandelte Kontrolle (190-901) nach 1:1 Paarbildung [57]

Endpunkt	Analysepopulation <sup>1</sup> 190-201	n 190-201	n 190-901	190-201 versus 190-901	
				HR (95% KI)	p-Wert <sup>2</sup>
Zeit bis zur ersten stabilen Abnahme des Scores um 2 Punkte oder Score = 0	ITT (Primäre Analyse)	22	22	0,10 (0,03; 0,38)	0,0005

<sup>1</sup> ITT bezieht sich auf die Studie 190-201

<sup>2</sup> Cox-Modell adjustiert nach Alter, ML/HML-Score zu Baseline.

Kriterien zur Paarbildung (matching): Identischer ML/HML-Skalenwert und Altersunterschied ≤12 Monate zu Baseline (300 mg-Baseline in Studie 190-201).

Die Daten wurden mit einem Cox-Modell (adjustiert nach ML/HML-Score und Alter jeweils zu Baseline) ausgewertet.

Eine stabile Abnahme entspricht einer Abnahme des Scores, der bis zum letzten Untersuchungszeitpunkt nicht wieder auf den Baseline-Wert (oder höher) angestiegen war. Ein stabiler Score von Null entspricht einer Abnahme auf 0 Punkte, die bis zum letzten Untersuchungszeitpunkt stabil bleibt.

ISE: Abbildung 3.2.2

In Abbildung 14 sind die Kaplan-Meier-Kurven der Zeit, bis sich der ML/HML-Score nachhaltig um zwei Punkte verschlechtert hatte oder nachhaltig bei null lag, bei den behandelten (190-201/202) bzw. unbehandelten Patienten (190-901) dargestellt (Woche 49). Die Abbildung zeigt, dass die Krankheitsprogression (nach der oben angegebenen Definition) bei den behandelten Patienten (300 mg-Dosisintervall der Studie 190-201/202), verglichen mit den unbehandelten Patienten (190-901), „dramatisch“ verzögert ist. Die Patienten der historischen Kontrolle (Studie 190-901) hatten zu jedem Zeitpunkt im Verlauf der Studie ein 10-fach höheres Risiko einer Krankheitsprogression (gemäß obiger Definition) als die behandelten Patienten (Studie 190-201/202). Die mediane Zeit bis zur Krankheitsprogression (nach obiger Definition) betrug bei den Patienten der historischen Kontrolle 285 Tage (95% KI: 210; 420), während der Median bei den behandelten Patienten noch nicht erreicht ist.

Die Ergebnisse des 96-Wochen Follow-up sind in Kapitel 4.4.2.3.1.6 dargestellt.

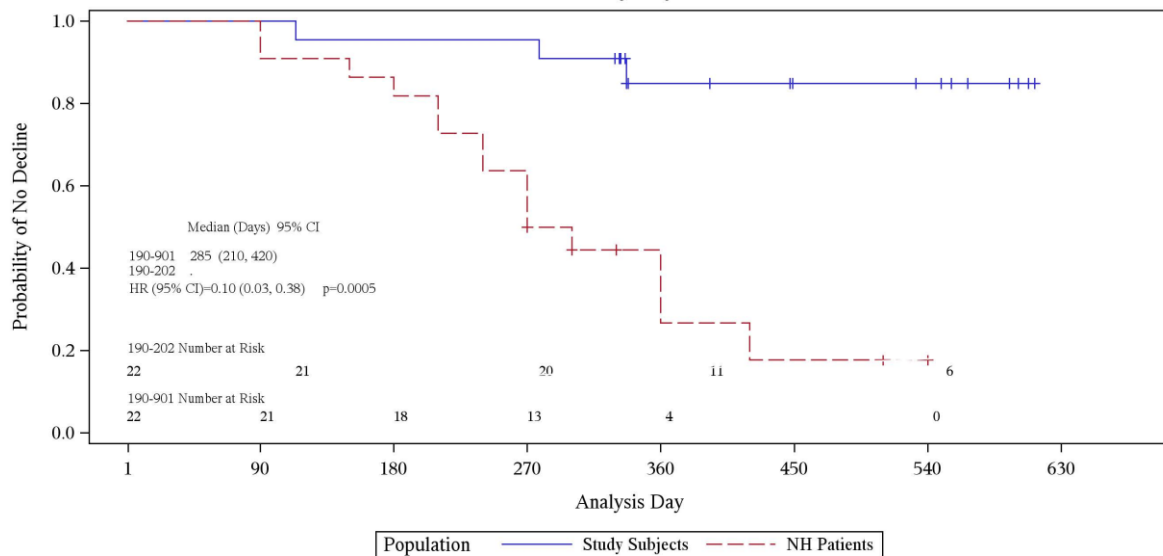


Abbildung 14: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Abnahme um  $\geq 2$  Punkte oder Wert von Null): Cerliponase alfa (190-201/202; ITT) vs. unbehandelte Kontrolle (190-901) nach 1:1 Paarbildung (Alter, ML/HML Wert) [57]

NH natural history, CI confidence interval, HR hazard ratio

Zeit bis zur ersten stabilen Abnahme des ML/HML-Scores um zwei Punkte von Baseline bis zum letzten Erhebungszeitpunkt oder Wert von Null. Eine stabile Abnahme entspricht einer Abnahme des Scores um zwei Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder auf eine Abnahme um einen Punkt (oder höher) angestiegen war. Ein Score von Null entspricht einer Abnahme auf 0 Punkte, die bis zum letzten Untersuchungszeitpunkt stabil blieb.

Die Auswertung zum 49-Wochen Zeitpunkt berücksichtigt CLN2 Daten bis zum Tag 340 relativ zur ersten 300 mg Infusion für 190-201/202 Patienten und relativ zum Alters-Matching für 190-901 Patienten. Zu Woche 49 waren 12 190-201/202 Patienten und 5 190-901 Patienten unter Risiko.

Kriterien zur Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq 12$  Monate zu Baseline (300 mg-Baseline in Studie 190-201).

Analysepopulation: ITT (n=22 nach Paarbildung)

Die Daten wurden in einem Kaplan-Meier- und Cox-Modell (adjustiert nach ML/HML-Score und Alter jeweils zu Baseline) analysiert.

ISE: Abbildung 3.2.2

### **Sensitivitätsanalysen (Ereigniszeitanalysen)**

Für diesen Endpunkt (Zeit bis zum ersten nachhaltigen Abfall um zwei Punkte oder ML/HML-Score = 0) wurden zusätzliche Auswertungen nach Kaplan-Meier als Sensitivitätsanalysen durchgeführt:

- Ausschluss von zwei Patienten der Studie 190-201/202 mit einem ML-Score von sechs zu Baseline, der im Studienverlauf nicht abfiel (n=2) (HR: 0,10; p=0,0005) (ISE Abbildung 3.2.3; ISE Tabelle 3.2.3.2)
- Ausweitung des Algorithmus zur Paarbildung hinsichtlich der Altersdifferenz der Paare auf 21 Monate (n=23) (HR 0,10, p=0,0003) (ISE Abbildung 3.2.4; ISE Tabelle 3.2.4.2);



- Adjustierung des Cox-Modells nach Alter, CLN2-Score und zusätzlich der Variable „Geschlecht“ (n=22) (HR 0,09, p=0,0005) (ISE Abbildung 3.2.10)

Bei allen Sensitivitätsanalysen zeigte sich übereinstimmend, dass die Patienten der Studie 190-901 ein 10-fach höheres Risiko hatten, dass die Erkrankung fortschreitet (gemäß obiger Definition).

Tabelle 4-47: Sensitivitätsanalysen der primären Kaplan-Meier Analyse mit 1:1 Paarbildung [57]

Endpunkt	Analysepopulation <sup>1</sup> 190-201	n 190-201	n 190-901	190-201 versus 190-901	
				HR (95% KI)	p-Wert <sup>5</sup>
Zeit bis zur ersten stabilen Abnahme des Scores um $\geq 2$ Punkte oder Score = 0	ITT (Primäre Analyse)	22	22	0,10 (0,03; 0,38)	0,0005
	ITT „Wirksamkeit“, ML <6 <sup>3</sup>	20	20	0,10 (0,03; 0,36)	0,0005
	ITT, 21 Monate <sup>2</sup>	23	23	0,10 (0,03; 0,35)	0,0003
	ITT <sup>4</sup>	22	22	0,09 (0,02; 0,35)	0,0005

<sup>1</sup> ITT bezieht sich auf die Studie 190-201  
<sup>2</sup> Erweiterung der 1:1 Paarbildungskriterien um Altersdifferenz von bis zu 21 Monaten  
<sup>3</sup> Ausschluss von zwei Patienten mit einem stabilen ML Score von 6.  
<sup>4</sup> Cox-Modell adjustiert nach Alter, Geschlecht und ML/HML-Score zu Baseline.  
<sup>5</sup> Die Daten wurden mit einem Cox-Modell (adjustiert nach ML/HML-Score und Alter jeweils zu Baseline) ausgewertet (wenn nicht anders angegeben).  
Kriterien zur Paarbildung (matching): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq 12$  Monate zu Baseline, wenn nicht anders angegeben (300 mg-Baseline in Studie 190-201).  
Eine stabile Abnahme entspricht einer Abnahme des Scores, der bis zum letzten Untersuchungszeitpunkt nicht wieder auf den Baseline-Wert (oder höher) angestiegen war. Ein Score von Null entspricht einer Abnahme auf 0 Punkte, die bis zum letzten Untersuchungszeitpunkt stabil blieb.  
ISE: Abbildung 3.2.2, Abbildung 3.2.3, Abbildung 3.2.4, Abbildung 3.2.10

In weitere Kaplan-Meier-Analysen wurden die gesamte ITT-Population der Studie 190-201/202 (ohne Paarbildung) und die auswertbaren Patienten der Studie 190-901 eingeschlossen. In Abbildung 15 sind diese Analysen bei den Patienten der Studie 190-201/202 in Form einer durchgezogenen Linie dargestellt und bei den Patienten der historischen Kontrolle in Form unterschiedlich gestrichelter Linien, abhängig von deren HML-Score zu Baseline (zwischen zwei und sechs Punkte). In diese Analyse wurden die auswertbaren Patienten der historischen Kontrolle (190-901; n=41) aufgenommen.

In Abbildung 15 ist die Zeit bis zur dauerhaften Verschlechterung um  $\geq 2$  Punkte unter Einsatz dieser Analysemethode dargestellt. Die Abbildung zeigt, dass der HML-Score bei allen Patienten der Studie 190-901 mit etwa der gleichen Rate abfällt, unabhängig von deren HML-Ausgangswert (Baseline; im Vergleich zu den behandelten Patienten der Studie 190-201/202, bei denen ein nachhaltiger Abfall um  $\geq 2$  Punkte nur selten auftrat)

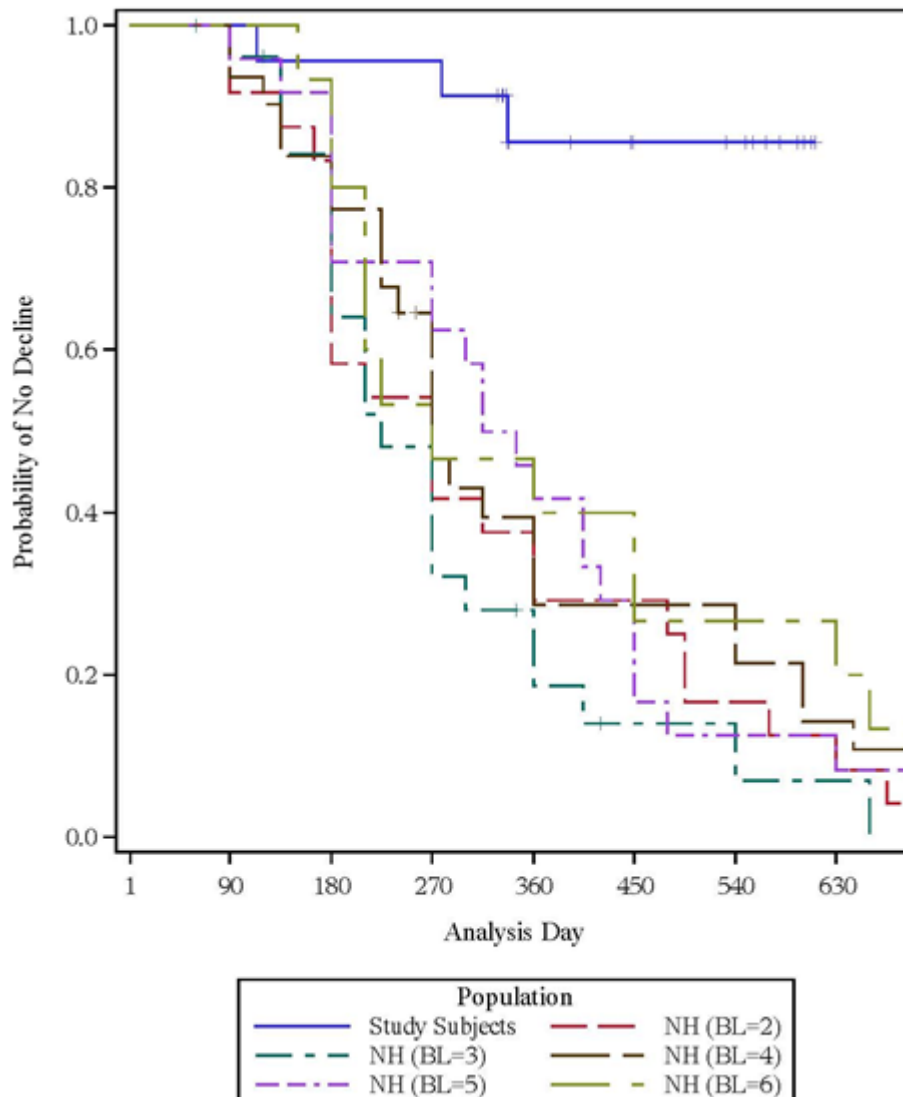


Abbildung 15: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Zeit bis zur ersten dauerhaften Abnahme um  $\geq 2$  Punkte): Cerliponase alfa (190-201/202; ITT) vs. unbehandelte Kontrolle (190-901; Auswertbare Patienten, n=41) ohne Paarbildung [57]

NH = Natural history, BL = HML-Score zu Baseline (Ausgangswert)

Studie 190-901: Zur Bestimmung des Ausgangswerts (Baseline) wurde der erste HML-Wert unter 6 ermittelt und bei aufeinander folgenden Werten unter 6 wurde der mittlere Wert zwischen dem ersten und letzten dieser Werte als Baseline definiert.

Quelle: ISE Abbildung 3.2.7

#### 4.4.2.3.1.4 Progressionsrate (Verschlechterung des ML/HML-Skalenwerts)

Die Rate, mit der der ML/HML-Score abfiel, war *a priori* als bedeutender Endpunkt festgelegt. Diese Ergebnisse werden unterstützend zur weiteren Information dargestellt, da dies der

ursprüngliche primäre Endpunkt war (später auf Empfehlung der Zulassungsbehörden geändert).

In die Analyse (Abfall des HML/ML-Scores in Punkten über 48 Wochen) wurden die ITT-Population der Studie 190-201/202 und die auswertbaren Patienten der Studie 190-901 nach 1:1 Paarbildung (Methode #1) eingeschlossen. Der HML/ML-Score (in Punkten/48 Wochen) fiel bei den Patienten der Studie 190-901 im Mittel (SD) um 2,06 (1,379) Punkte ab verglichen mit 0,53 (0,737) Punkten bei den Patienten der Studie 190-201/202, was einem Unterschied von 1,53 Punkten/48 Wochen ( $p < 0,0001$ ) entspricht. Die verschiedenen, unterstützenden und Sensitivitätsanalysen zeigten übereinstimmend einen ähnlichen und hoch signifikanten Unterschied zwischen den behandelten und unbehandelten Patienten.

Von Bedeutung ist, dass die Ergebnisse auch dann vergleichbar waren, wenn kein Algorithmus zur Paarbildung eingesetzt wurde (Tabelle 4-48).

Tabelle 4-48: Geschätzte Abnahme des ML/HML-Scores/48 Wochen: Cerliponase alfa (190-201) vs. unbehandelte Kontrolle (190-901) mit und ohne 1:1 Paarbildung [57]

Analyse	n	Abfall/48 Wochen 190-901 Mittelwert (SD)	Abfall/48 Wochen 190- 201/202 Mittelwert (SD)	Differenz (901 - 201/202) Mittelwert (SE)	95% KI der Differenz <sup>1</sup>	Zweiseitiger p-Wert <sup>1</sup>
<i>Analysen mit Paarbildung („matching“)</i>						
ITT, Methode #1 (prädefinierte Analyse)	22	2,06 (1,379)	0,53 (0,737)	1,53 (0,333)	0,85; 2,21	<0,0001
ITT, Methode #2 <sup>7</sup>	23	2,03 (1,359)	0,48 (0,756)	1,55 (0,324)	0,89; 2,21	<0,0001
ITT „Wirksamkeit“, ML <6 <sup>2</sup>	20	2,08 (1,390)	0,58 (0,754)	1,49 (0,353)	0,77; 2,22	0,0002
ITT, 21 Monate <sup>3</sup>	23	2,06 (1,348)	0,48 (0,756)	1,57 (0,322)	0,92; 2,23	<0,0001
„Wirksamkeit“, ML <6, ≥3 <sup>2,4</sup>	19	2,19 (1,336)	0,62 (0,761)	1,57 (0,353)	0,85; 2,29	0,0001
<i>Analysen ohne Paarbildung</i>						
„Wirksamkeit“, ML <6, Mittlerer Wert <sup>2,5</sup>	-	2,09 (0,966) (n=41)	0,53 (0,776) (n=21)	1,56 (0,243)	1,11; 2,02	<0,0001
„Wirksamkeit“, ML <6, ML ≥3, Mittlerer Wert <sup>2,5,6</sup>	-	2,09 (0,966) (n=41)	0,54 (0,828) (n=18)	1,55 (0,262)	1,05; 2,05	<0,0001
„Wirksamkeit“, ML <6, ML ≥2, Mittlerer Wert <sup>2,5,8</sup>	-	2,09 (0,966) (n=41)	0,56 (0,787) (n=20)	1,54 (0,249)	1,07; 2,00	<0,0001

Analyse	n	Abfall/48 Wochen 190-901 Mittelwert (SD)	Abfall/48 Wochen 190- 201/202 Mittelwert (SD)	Differenz (901 - 201/202) Mittelwert (SE)	95% KI der Differenz <sup>1</sup>	Zweiseitiger p-Wert <sup>1</sup>
„Wirksamkeit“, ML <6, Erster/letzter Wert <sup>2,9</sup>	-	1,95 (0,922) (n=41)	0,53 (0,776) (n=21)	1,42 (0,235)	0,98; 1,87	<0,0001
„Wirksamkeit“, ML <6 <sup>2,5</sup> , Regression <sup>10</sup>	-	2,09 (0,988) (n=41)	0,48 (0,795) (n=21)	1,61 (0,249)	1,14; 2,08	<0,0001
„Wirksamkeit“ ML <6 <sup>2,5</sup> , Slope/ Intercept <sup>11</sup>	-	1,98 (SE 0,137) (n=41)	0,44 (SE 0,187) (n=21)	1,53 (0,232)	1,08; 1,99	<0,0001

<sup>1</sup> Zweistichproben-t-Test mit Satterthwaite-Approximation. Zur Berechnung der KIs wurde eine t-Verteilung unter Einsatz der Satterthwaite-Approximation (KIs des Unterschieds) herangezogen.

<sup>2</sup> Ausschluss von zwei Patienten mit stabilem ML-Score von 6.

<sup>3</sup> Erweiterung der 1:1 Paarbildungskriterien um Altersdifferenz von bis zu 21 Monaten (post-hoc)

<sup>4</sup> Ausschluss eines Paares mit einem Baseline-ML-Score von 1.

<sup>5</sup> 190-901: Baseline: Mittlerer Wert einer Reihe aufeinander folgender Werte ( $\leq 5$ ) nach dem ersten Wert <6.

<sup>6</sup> Ausschluss eines Patienten (190-201) mit Baseline-ML-Score <3.

<sup>7</sup> Siehe Tabelle 4-75

<sup>8</sup> Ausschluss von Patienten (190-201) mit einem 300 mg-ML-Score = 1 zu Baseline (post-hoc).

<sup>9</sup> Baseline 190-901: Frühester ML-Wert <6

<sup>10</sup> Lineare Regression zur Schätzung der Intra-Subject-Rate des Abfalls (Slope); ML/HML-Score als lineare Funktion der Zeit

<sup>11</sup> Random-Intercept/Slope-Modell (MMRM). Faktoren: Studie, Alter zu Baseline, ML/HML-Score zu Baseline, Erhebungszeitpunkt, Interaktion Studie/Zeit; unstrukturierte Kovarianz-Matrix

Positive Werte entsprechen einer Krankheitsprogression und negative Werte einer Verbesserung des klinischen Zustands. Die Rate wurde bei jedem Patienten (Intra-Subjekt) aus der Neigung (Slope) der Linie zwischen zwei ML/HML-Erhebungen (Phase der klinischen Progression) geschätzt (wenn nicht anders angegeben).

Kriterien zur Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq 12$  Monate zu Baseline wenn nicht anders angegeben (300 mg-Baseline in Studie 190-201).

ISE: Tabelle 1.1.2, Tabelle 1.1.3, Tabelle 1.1.3.2, Tabelle 1.1.4, Tabelle 1.1.6, Tabelle 1.1.7, Tabelle 1.1.7.2, Tabelle 1.1.11, Tabelle 1.1.9, Tabelle 1.1.11, Tabelle 1.1.12

#### 4.4.2.3.1.5 Einzeldomänen der ML/HML-Skala und andere Summenscores

##### *Einzeldomänen: Motorische und Sprach-Subskala*

Wie bereits bei der Analyse der kombinierten ML/HML-Skala, zeigte sich auch hinsichtlich des Abfalls der Einzeldomänen (motorisch und Sprache getrennt) ein ähnlicher Behandlungseffekt zugunsten der behandelten Patienten der Studie 190-201-202. Der mittlere Gruppenunterschied (SE) betrug hinsichtlich der motorischen Funktion 0,72 (0,240) Punkte/ 48 Wochen ( $p=0,0055$ ) und 1,01 (0,200) Punkte/48 Wochen ( $p<0,0001$ ; Tabelle 4-49) hinsichtlich der sprachlichen Fähigkeiten.

Tabelle 4-49: Geschätzte Abnahme der motorischen bzw. Sprachdomäne (jeweils 0-3 Punkte)/48 Wochen: Cerliponase alfa (190-201, ITT) vs. unbehandelte Kontrolle (190-901) mit 1:1 Paarbildung [57]

Analyse	n	Abfall/48 Wochen 190-901 Mittelwert (SD)	Abfall/48 Wochen 190-201/202 Mittelwert (SD)	Differenz (901 – 201/202) (SE)	95% KI der Differenz <sup>1</sup>	Zweiseitiger p-Wert <sup>1</sup>
Motorische Domäne	22	1,01 (1,046)	0,29 (0,415)	0,72 (0,240)	0,23; 1,21	0,0055
Sprachdomäne	21 <sup>2</sup>	1,10 (0,814)	0,09 (0,419)	1,10 (0,200)	0,61; 1,42	<0,0001

<sup>1</sup> Zweistichproben-t-Test mit Satterthwaite-Approximation. Zur Berechnung der KIs wurde eine t-Verteilung unter Einsatz der Satterthwaite-Approximation (KIs des Unterschieds) herangezogen.

<sup>2</sup> Ein Paar wurde ausgeschlossen, da bei mindestens einem Patienten des Paares der Baseline-Score = 0 war. Positive Werte entsprechen einer Krankheitsprogression und negative Werte einer Verbesserung des klinischen Zustands. Die Rate wurde bei jedem Patienten (Intra-Subjekt) aus der Neigung (Slope) der Linie zwischen zwei ML/HML-Erhebungen (Phase der klinischen Progression) geschätzt.

Kriterien zur Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied ≤12 Monate zu Baseline (300 mg-Baseline in Studie 190-201).

ISE: Tabelle 1.3.1, Tabelle 1.3.2

### ***Summenscore (motorische Funktion / Sprache / Sehvermögen)***

Zusätzlich wurde der Summenscore, bestehend aus den Domänen motorische Funktion, Sprache und Sehen, analysiert bzw. dessen Verlauf (MLV/HMLV-Score: 0-9 Punkte). Bei zwei der insgesamt 41 auswertbaren Patienten der Studie 190-901 lagen keine Angaben zur Subskala Visusverlust vor, sodass in diese Analyse 39 Patienten eingeschlossen wurden. Bei den auswertbaren Patienten der Studie 190-901 (n=39) nahm der MLV/HMLV-Score im Mittel (SD; berechnet als „Slope“) um 2,23 (1,152) Punkte/48 Wochen ab verglichen mit 0,78 (1,042) Punkten/48 Wochen bei den Patienten der Studie 190-201/202 (n=21, „Wirksamkeits“-Population). Der Gruppenunterschied betrug 1,44 Punkte/48 Wochen (p<0,0001) (Tabelle 4-50).

Tabelle 4-50: Geschätzte Abnahme des MLV/HMLV-Scores (0-9 Punkte)/48 Wochen: Cerliponase alfa (190-201) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) ohne Paarbildung [57]

Analyse	Abfall/48 Wochen 190-901 Mittelwert (SD) n=39 <sup>1</sup>	Abfall/48 Wochen 190-201/202 Mittelwert (SD) n=21 <sup>2</sup>	Differenz (901 - 201/202) Mittelwert (SE)	95% KI des Unterschieds <sup>4</sup>	Zweiseitiger p-Wert <sup>4</sup>
HMLV/MLV-Score <sup>3,7</sup>	2,23 (1,152)	0,78 (1,042)	1,44 (0,302)	0,85; 2,03	<0,0001
HMLV/MLV, lineare Regression <sup>5</sup>	2,30 (1,162)	0,74 (1,018)	1,56 (0,302)	0,98; 2,14	<0,0001
HMLV/MLV, Random-Slope/ Intercept <sup>6</sup>	2,13 (0,158) (SE)	0,63 (0,233) (SE)	1,49 (0,281)	0,94; 2,04	<0,0001

<sup>1</sup> Bei zwei Patienten lagen keine Werte zur Sehfähigkeit vor.  
<sup>2</sup> Ausschluss von Patienten mit einem stabilen ML-Score von 6.  
<sup>3</sup> 190-901: Baseline: Mittlerer Wert einer Reihe aufeinander folgender CLN2 Werte nach dem ersten Wert <9.  
<sup>4</sup> Zweistichproben-t-Test mit Satterthwaite-Approximation. Zur Berechnung der KIs wurde eine t-Verteilung unter Einsatz der Satterthwaite-Approximation (KIs des Unterschieds) herangezogen.  
<sup>5</sup> Lineare Regression zur Schätzung der Intra-Subjekt-Rate des Abfalls (Slope); CLN2-Score als lineare Funktion der Zeit  
<sup>6</sup> Random-Intercept/Slope-Modell (MMRM). Berücksichtigte Faktoren: Studie, Alter und CLN2-Score zu Baseline, Erhebungszeitpunkt, Interaktion Studie/Zeit; unstrukturierte Kovarianz-Matrix  
<sup>7</sup> Die Rate wurde bei jedem Patienten (Intra-Subjekt) aus der Neigung (Slope) der Linie zwischen zwei ML/HML-Erhebungen (Phase der klinischen Progression) geschätzt.  
Positive Werte entsprechen einer Krankheitsprogression und negative Werte einer Verbesserung des klinischen Zustands.  
ISE: Tabelle 1.2.1, Tabelle 1.2.2, Tabelle 1.2.3

### **Summenscore (motorische Funktion / Sprache / Sehvermögen / Krampfanfälle)**

Der mittlere Ausgangswert (SD) des Summenscores, gebildet aus der MLV-Skala plus Krampfanfälle (0-12 Punkte), betrug bei den Patienten der Studie 190-901 9,1 (1,79) Punkte (n=39) und bei den Teilnehmern der Studie 190-201/202 8,0 (1,83) Punkte (n=23). Bei den unbehandelten Patienten (190-901) wurde der Ausgangswert durch Interpolation berechnet (mittlerer Wert aus der ersten Reihe von Werten <12 Punkte) und für die Studie 190-201/202 wurde die 300 mg-Baseline als Ausgangswert herangezogen. Bei den 37 Patienten der Studie 190-901 betrug der Mittelwert (SD) in Woche 49 7,0 (2,41) Punkte und bei den 23 Patienten der Studie 190-201/202 7,9 (2,08) Punkte. Nach 48 Wochen hatte der Summenscore (Mittelwert  $\pm$  SD) bei den unbehandelten Patienten um 2,2 (1,85) Punkte abgenommen verglichen mit -0.1 (1.94) Punkten bei den behandelten Patienten (Tabelle 4-51).

Tabelle 4-51: Geschätzte Abnahme des Gesamtscores (0-12 Punkte)/48 Wochen: Cerliponase alfa (190-201/202 ITT, 300 mg-Dosisintervall) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) ohne Paarbildung [57]

Gruppe	Mittelwert ± SD (Punkte) <sup>1</sup>				
	Baseline	Woche 25	Woche 49	Änderung bis Woche 49	Änderung in Woche 73
190-901 (n=39)	9,1 (1,79)	-1,3 (1,31)	7,0 (2,41) (n=37)	-2,2 (1,85) (n=37)	-3,7 (1,95) (n=34)
190-201/202 (n=23)	8,0 (1,83)	-0,3 (1,62)	7,9 (2,08)	-0,1 (1,94)	1,3 (1,33) (n=9)

<sup>1</sup> Geschätzt durch Interpolation.  
ISE: Tabelle 2.3.1

#### 4.4.2.3.1.6 Ergebnisse der weiteren Nachverfolgung

Zur Nachverfolgung von Wirksamkeit und Sicherheit führte BioMarin weitere Analysen mit Datenschnitt vom 03.06.2016 [34,66] oder 01.11.2016 [59–62] durch. Zu diesen Zeitpunkten waren alle Patienten über mindestens 72 bzw. 96 Wochen behandelt worden (ausgenommen ein Patient, der die Studie nach nur einer Dosis in der Titrationsphase abgebrochen hatte). Zum Follow-up nach 96 Wochen liegen vergleichende Ergebnisse vor, die im Folgenden dargestellt sind.

Die Ansprechrate, d. h. Anteil der Patienten im Gruppenvergleich, deren ML/HML-Score sich um weniger als zwei Punkte/48 Wochen verschlechtert hatte, war *a priori* als primärer Endpunkt bzw. Primäranalyse definiert (ISE). Zur Follow-up-Analyse des primären Endpunkts wurden aktualisierte, klinische Daten herangezogen, sodass alle Patienten über mindestens 96 Wochen behandelt worden waren. Um die fehlende Randomisierung auszugleichen, wurde eine 1:1 Paarbildung („matching“) eingesetzt; hierbei wurden nur Paare gebildet, deren Alter sich zu Baseline um 12 Monate oder weniger unterschied und deren ML-Score zu Baseline exakt übereinstimmte (Baseline entspricht dem Wert vor der ersten Dosis von 300 mg). Diese Kriterien trafen auf 21 Paare zu, die aus der Studienpopulation (n=23) und der historischen Kontrolle (n=42) gebildet wurden. Von diesen Paaren sprachen alle der 21 Patienten der Studienpopulation auf die Behandlung an (Abfall um <2 Punkte/48 Wochen), während nur 9 Patienten der historischen Kontrolle (43%) das Responder-Kriterium erfüllten. Der in der Primäranalyse gemessene Behandlungseffekt war dauerhaft und blieb im Studienverlauf unverändert erhalten (Tabelle 4-52).

Tabelle 4-52: Ansprechrate gemessen mittels ML/HML-Skala (<2 oder ≥2 Punkte/48 Wochen): Cerliponase alfa (190-201; ITT) vs. unbehandelte Kontrolle (190-901, auswertbare Patienten) nach 1:1 Paarbildung; DCO: 01 Nov 2016 [59]

Response	Studie		Differenz (190-201) - (190-901)	p-Wert <sup>1</sup> Differenz	190-201/190-901		
	190-901 (n=21)	190- 201/202 (n=21)			RR	95% KI	p-Wert
Ja (Rate <2)	9 (43%)	21 (100%)	57%	<0,0001	2,2632	1,3955; 3,6704	0,0009
Nein (Rate ≥2)	12 (57%)	0	-	-	0,0400	0,0025; 0,6347	0,0225

<sup>1</sup> Exakter Test nach Fisher  
 KI: Konfidenz Intervall; RR: relatives Risiko  
 Kriterien zur 1:1 Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied ≤12 Monate zu Baseline (300 mg-Baseline in Studie 190-201). Bei Patienten der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1 Paarbildung.  
 Geschätzte Neigung (Slope; Intra-Subjekt) kategorisiert nach <2 und ≥2 Punkte auf der ML/HML-Skala  
 Quelle: Tabelle 3.1.1b [59]

Zur Schätzung der Zeit bis zur nachhaltigen Verschlechterung des ML-Scores um ≥2 Punkte oder auf den Wert von Null (ML/HML = 0) wurde die Kaplan-Meier-Methode eingesetzt. Zusätzlich wurden die Daten mittels eines Cox-proportionalen Hazard-Modells ausgewertet, welches das Alter und den ML-Score zu Baseline (erste 300 mg-Dosis) als Kovariaten enthielt. In dem Cox-Modell wurde die gesamte Nachbeobachtungszeit analysiert und keine Imputation eingesetzt, da die Methoden den variablen Zeitplan und die Dauer der Nachbeobachtung bereits berücksichtigen. Kriterien zur Paarbildung waren der ML-Score und das Alter (innerhalb von 12 Monaten), jeweils zu Baseline. Der Unterschied der Ereignisraten zwischen BMN190 und den unbehandelten Patienten (190-901) nahm mit der Dauer der Behandlung zu. Bei den Patienten der historischen Kontrolle betrug der Median, bis der HML-Score nachhaltig um zwei Punkte abgefallen war, 270 Tage, während der Median bei den behandelten Patienten nie erreicht wurde, da der Behandlungseffekt im Studienverlauf stabil blieb. Der Unterschied der Ereignisraten war zwischen den Gruppen in Woche 48 hoch signifikant und stieg bis Woche 72 und 96 weiter an. Der Gruppenunterschied, berechnet mittels eines Cox-Modells mit den Kovariaten Alter und ML-Score zu Baseline, war in Woche 96 hoch signifikant: HR = 0,05 ([95% KI 0,01; 0,17], p<0,0001). Wie die Primäranalyse der ML-Skala zeigt, blieb der „dramatische“ Behandlungseffekt durchgehend bis Woche 96 erhalten (Abbildung 16) [59].



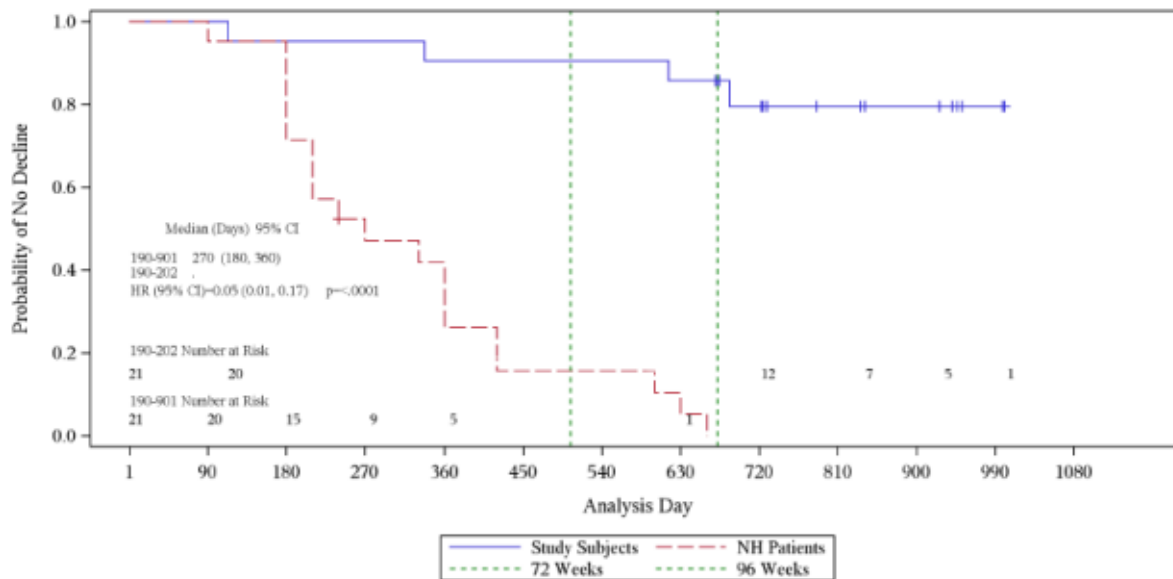


Abbildung 16: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Abnahme um  $\geq 2$  Punkte oder Wert von Null): Cerliponase alfa (190-201/202) vs. unbehandelte Kontrolle (190-901) mit 1:1 Paarbildung (Alter, ML/HML Wert); DCO: 01 Nov 2016 [59]

NH natural history, CI confidence interval, HR hazard ratio

Zeit bis zur ersten stabilen Abnahme des ML/HML-Scores um zwei Punkte von Baseline bis zum letzten Erhebungszeitpunkt oder Wert von Null. Eine stabile Abnahme entspricht einer Abnahme des Scores um zwei Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder auf eine Abnahme um einen Punkt (oder höher) angestiegen war. Ein Score von Null entspricht einer Abnahme auf 0 Punkte, die bis zum letzten Untersuchungszeitpunkt stabil blieb.

Kriterien zur Paarbildung („matching“): Identischer ML/HML-Skalenwert und Altersunterschied  $\leq 12$  Monate zu Baseline (300 mg-Baseline in Studie 190-201).

Konservative Analysepopulation: von den in Studie 190-201 eingeschlossenen Patienten wurden zwei Patienten mit einem stabilen ML Score von 6 ausgeschlossen. Der Studienabbruch eines Patienten, der nur eine Dosis erhalten hatte und aus der ITT Population ausgeschlossen war, wurde als Event erhoben.

Die Daten wurden in einem Kaplan-Meier- und Cox-Modell (adjustiert nach ML/HML-Score und Alter jeweils zu Baseline) analysiert.

Quelle: Abb. 3.2.1b [59]

Bei einer weiteren Analyse wurde die Zahl der Mutationen als zusätzliches Kriterium zur Paarbildung herangezogen (exakte Übereinstimmung hinsichtlich der Zahl der Hauptmutationen), und der Altersunterschied zu Baseline wurde auf höchstens drei Monate beschränkt. Wie bei der Analyse des primären Endpunkts, wurden zwei Patienten ausgeschlossen, deren Screening ML-Score von sechs während der Studie stabil bei ML = 6 blieb. Der Patient, der die Studie frühzeitig abgebrochen hatte, wurde als Non-Responder gewertet. Abbildung 17 zeigt die Kaplan-Meier-Kurven der beiden Gruppen und bestätigt den großen Behandlungseffekt in Woche 48, der bis Woche 72 und 96 noch zunimmt. Die geschätzte HR betrug 0,11 ([95% KI 0,03; 0,37],  $p=0,0003$ ).

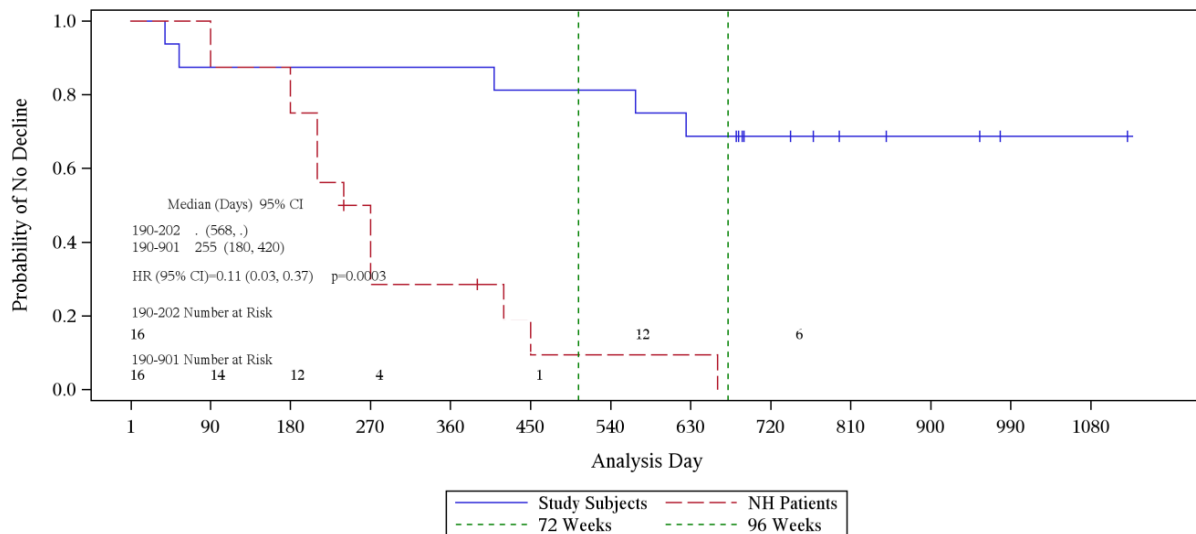


Abbildung 17: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML/HML-Skala (Abnahme um  $\geq 2$  Punkte oder Wert von Null): Cerliponase alfa (190-201/202) vs. unbehandelte Kontrolle (190-901) mit 1:1 Paarbildung (Alter, Mutation, ML/HML Wert); DCO: 01 Nov 2016 [59]

NH natural history, CI confidence interval, HR hazard ratio

Zeit bis zur ersten stabilen Abnahme des ML/HML-Scores um zwei Punkte von Baseline bis zum letzten Erhebungszeitpunkt oder Wert von Null. Eine stabile Abnahme entspricht einer Abnahme des Scores um zwei Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder auf eine Abnahme um einen Punkt (oder höher) angestiegen war. Ein stabiler Score von Null entspricht einer Abnahme auf 0 Punkte, die bis zum letzten Untersuchungszeitpunkt stabil blieb.

Kriterien zur Paarbildung („matching“): Identischer ML/HML-Skalenwert, Altersunterschied  $\leq 3$  Monate, Anzahl der identischen Hauptmutationen zu Baseline (Screening-Baseline in Studie 190-201).

Konservative Analysepopulation: von allen in Studie 190-201 eingeschlossenen Patienten wurden zwei Patienten mit einem stabilen ML Score von 6 ausgeschlossen. Der Studienabbruch eines Patienten, der nur eine Dosis erhalten hatte und aus der ITT Population ausgeschlossen war, wurde als Event erhoben.

Die Daten wurden in einem Kaplan-Meier- und Cox-Modell (adjustiert nach ML/HML-Score und Alter jeweils zu Baseline) analysiert.

Quelle: Abb. 2.1.2.5 [59]

### **Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext**

Die „Hamburg“-CLN2-Skala wurde erstmals von Steinfeld et al. im Jahr 2002 beschrieben [5], wurde allerdings, wie die Studien zum natürlichen Verlauf der Erkrankung belegen, schon viele Jahre zuvor in Hamburg eingesetzt. Mit den Änderungen der ML-Skala (die in den Cerliponase alfa-Studien eingesetzt wird) wurde den Empfehlungen von CLN2-Experten unter Leitung eines erfahrenen Sachverständigen entsprochen. Die Änderungen betrafen Ankerpunkte, um zu erreichen, dass die Bewertungen im Rahmen einer multizentrischen Studie sowohl zwischen den einzelnen Prüfarzten als auch über die Zeit hinweg konsistent sind. Die Änderungen waren nur geringfügig, der eigentliche Inhalt blieb unverändert, und jede Stufe (respektive jeder Punkt auf der Skala) entspricht einem „Meilenstein“ in der kindlichen Entwicklung. Die verblindete Brückenstudie stellt sicher, dass die HML-Skala (eingesetzt in den Studien zum natürlichen

Verlauf) und die ML-Skala vergleichbar sind [24]. In Deutschland ist die HML-Skala Teil der klinischen Praxis, um die Schwere der CLN2 zu bewerten.

Das deutsche Zentrum war das größte Zentrum der Studie 190-201/202, und die Hälfte der Patienten wurden in diesem Zentrum eingeschlossen (n=12 von 24 eingeschlossenen Patienten).

Der Vergleich der Studien 190-201/202 und 190-901 zeigt, dass die Studienpopulation (190-201/202) eine repräsentative Stichprobe der gesamten Patientenpopulation darstellt (Abbildung 8). Der Behandlungseffekt (behandelte vs. unbehandelte Patienten) ist zwischen den Analysen mit und ohne Paarbildung vergleichbar (Tabelle 4-48). Die meisten Patienten der historischen Kontrolle wurden in Hamburg rekrutiert: n=33/41 auswertbaren Patienten.

Bei allen Studienzentren handelt es sich um qualifizierte CLN2-Zentren in westlichen Ländern, und es ist unwahrscheinlich, dass sich die medizinischen Standards (einschließlich symptomatische Behandlung) zwischen den Zentren wesentlich unterscheiden, obwohl es keine klar definierten Therapiestandards gibt. Da es während der letzten Jahrzehnte keine bedeutenden Neuerungen bei der Behandlung der CLN2 gegeben hat, ist nicht anzunehmen, dass sich die symptomatische Behandlung oder andere therapeutische Maßnahmen wesentlich zwischen den Patienten der Studie 190-201/202 und den Patienten der historischen Kontrolle unterscheiden. Allerdings liegen bei den Patienten der historischen Kontrolle keine entsprechenden Daten vor. Alle derzeit zur Behandlung der CLN2 eingesetzten Arzneimittel wirken rein symptomatisch und verändern den Krankheitsverlauf nicht, d. h. Begleitmedikationen reichen allein nicht aus, um den erheblichen Effekt von Cerliponase alfa auf den Abfall des ML-Scores plausibel zu erklären [40]. Darüber hinaus gab es in der Studie 190-201 keine Beschränkungen hinsichtlich der Begleitmedikation (abgesehen davon, dass diese, soweit möglich, nicht verändert werden sollte) oder anderer Therapiemaßnahmen. Daher wird nicht angenommen, dass sich die Art der Begleitbehandlung zwischen der Studie und der in Deutschland üblichen Praxis unterscheidet (insbesondere in Anbetracht der Tatsache, dass das deutsche Zentrum am größten war).

Die Ergebnisse aus den Erhebungen der ML-Skala, der Einzeldomänen sowie der MLV- und Gesamt-Skala sind auf den deutschen Versorgungskontext übertragbar. Sollte es hinsichtlich der Versorgungsstandards geringfügige Unterschiede zwischen der Studienpopulation und den Patienten der historischen Kontrolle geben, können diese nicht den erheblichen Behandlungseffekt von Cerliponase alfa erklären.

#### 4.4.2.3.2 Lebensqualität gemessen mittels PedsQL™ und CLN2 spezifischem Fragebogen - weitere Untersuchungen

##### PedsQL™ (Module: „Parent report for toddlers“ und „Parent family impact“)

Tabelle 4-53: Operationalisierung des Endpunkts gesundheitsbezogene Lebensqualität (PedsQL™) – weitere Untersuchungen

Studie	Operationalisierung
190-201/202	Die Erhebung und Dokumentation beider PedsQL™-Fragebögen („Parent Report for Toddlers“ und „Family Impact Module“) entsprachen den Vorgaben von James W. Varni, PhD, der das

Instrument entwickelt hat: “Scaling and scoring of the Pediatric Quality of Life Inventory PedsQL™”. Bei beiden Fragebögen handelt es sich um generische Instrumente, die von Angehörigen beantwortet wurden [50]. Die PedsQL™-Fragebögen sind formell nicht für CLN2-Patienten validiert.

Der “Parent Report for Toddlers” setzt sich aus 21 Items, die vier Dimensionen abdecken, zusammen:

- Körperliche Kompetenzen (8 Items: Probleme beim Gehen, Laufen, der aktiven Teilnahme an Spiel oder Sport, Heben von etwas Schwerem, Baden, Helfen, seine oder ihre Spielwaren aufzuheben, Verletzungen oder Schmerzen, niedriges Energieniveau)
- Emotionale Kompetenzen (5 Items: Angstgefühle, Traurigkeit oder Depressionen, Schlafprobleme und Sorgen)
- Soziale Kompetenzen (5 Items: Probleme, mit anderen Kindern zu spielen, andere Kinder wollen nicht mit ihr oder ihm spielen, von anderen Kindern geärgert zu werden, nicht in der Lage sein, Dinge zu tun, die andere Kinder ihres/seines Alters können, beim Spiel mit anderen Kindern mitzuhalten)
- Schulische Fähigkeiten (3 Items: Probleme, in der Schule mit den Schulkameraden mitzuhalten, Fehltage in der Schule oder dem Kindergarten aufgrund von Unwohlsein oder Arztbesuchen/Krankenhausaufenthalten).

Jedes Item wurde auf einer 5-Punkt Likertskala von 0 (nie) bis 4 (fast immer) bewertet. Die Werte wurden wie folgt in eine Skala von 0-100 umgewandelt: 0=100, 1=75, 2=50, 3=25, 4=0.

Aus der Summe der beantworteten Fragen wurde der Mittelwert gebildet.

Psychosozialer Gesundheits-Summenscore (“Psychosocial Health Summary Score”): Summe aus den beantworteten Items der Dimensionen emotionale und soziale Kompetenzen sowie Fähigkeiten in der Schule.

Gesamtscore: Summe der beantworteten Items aus allen Dimensionen.

Das PedsQL™ “Family Impact Module” besteht aus 36 Items, die 8 Dimensionen abdecken. In der Studie 190-201 wurden fünf der 8 Dimensionen erfasst:

- Körperliche Kompetenzen (6 Items: Müdigkeit, Kopfschmerzen, Schwächegefühle und Magenprobleme)
- Emotionale Kompetenzen (5 Items: Angst, Traurigkeit, Wut, Frustration oder Gefühl, hilflos und hoffnungslos zu sein)
- Soziale Kompetenzen (4 Items: Gefühl der Isolierung, Schwierigkeiten, Hilfe von anderen zu erhalten und Zeit und Energie für soziale Aktivitäten zu finden)
- Kognitive Kompetenzen (5 Items: Kognitive Probleme einschließlich Aufmerksamkeit oder sich an Dinge zu erinnern und schnell zu denken)
- Kommunikation (3 Items: Probleme mit Anderen, die die familiäre Situation nicht verstehen, Schwierigkeiten, über den Gesundheitszustand des Kindes zu sprechen und mit Gesundheitsfachleuten zu kommunizieren)

Die Dimensionen Sorgen, tägliche Aktivitäten und Beziehungen innerhalb der Familie wurden nicht erhoben [50]. Jedes Item wurde wie oben beschrieben auf einer 5-Punkt Likertskala bewertet und in einen Score von 0-100 rückverwandelt.

Aus der Summe der beantworteten Items wurde der Mittelwert gebildet.

Summenscore der gesundheitsbezogenen Lebensqualität der Eltern (20 Items): Summe aus der Zahl der beantworteten Items aus den Dimensionen körperliche, emotionale, soziale und kognitive Fähigkeiten.

(Modifizierter) Summenscore: Summe aus der Zahl der über alle Dimensionen beantworteten Items.

Beide Skalen wurden nicht ausgewertet, wenn mehr als 50% der Items der Skala fehlten. Dieses Vorgehen entspricht den Vorgaben des PedsQL™ “Scaling and scoring of the Pediatric Quality of Life Inventory PedsQL™” [50] sowie Publikationen zum PedsQL™ [49].

Daten zur Lebensqualität wurden vor der ersten Cerliponase alfa-Infusion (Baseline), zu Beginn der stabilen Dosisphase und danach alle 12 Wochen sowie zu Studienende oder bei vorzeitigem

	<p>Abbruch erhoben. Die Fragebögen sollten möglichst vor den MRIs und den Probenahmen ausgefüllt werden. Baseline stellte die letzte Beobachtung vor der ersten 300 mg-Infusion dar; lag diese nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300 mg-Infusion eingegeben.</p> <p>Eine Imputation für fehlende Werte war nicht vorgesehen.</p> <p>Die berichteten Daten beziehen sich auf den finalen Datenschnitt der Studie 190-201.</p>
190-202	<p>Der Endpunkt wurde wie oben beschrieben operationalisiert, allerdings fanden die Erhebungen während der Behandlung alle 24 Wochen und zu Behandlungsende statt. Die Ergebnisse des Endpunkts werden nicht berichtet, da lediglich eine Zwischenanalyse der fortlaufenden Studie vorliegt und zu diesem Zeitpunkt nicht alle Patienten in die Studie eingeschlossen worden waren.</p>
190-901 (natural history)	<p>Der Endpunkt wurde in der Studie 190-901 nicht erhoben.</p>

Für Patienten der historischen Kontrollgruppe („matched“) liegen keine Daten zur Lebensqualität vor, und es können keine vergleichenden Analysen zwischen der Studie 190-201 und der historischen Kontrolle durchgeführt werden. Aus diesem Grund werden in diesem Dossier longitudinale Daten aus der klinischen Studie berichtet.

Der Mittelwert (SD) des PedsQL™-Moduls „Parent Report for Toddlers“ nahm zwischen Baseline und Woche 49 der stabilen Dosisphase um 2,6 (12,16) Punkte zu, d. h. der Wert verbesserte sich um etwa 4,3% (ITT, gesamtes Dosisintervall).

Fünf der insgesamt 8 Dimensionen des PedsQL™-Moduls „Family Impact“ wurden in der klinischen Studie erhoben. Der Mittelwert (SD) des Moduls „Family Impact“ stieg zwischen Baseline und Woche 49 der stabilen Dosisphase um 3,7 (19,04) Punkte an, d. h. verbesserte sich um circa 6,0%. Insgesamt änderten sich die einzelnen Domänen des PedsQL™ im Studienverlauf von der ersten (Baseline) bis zur letzten Beobachtung nur wenig. Nur die Werte zur körperlichen Leistungsfähigkeit (Domäne „physical function“; „Parent Report for Toddlers“) und gesellschaftlichen Funktion (Domäne „social function“; „Family Impact“) nahmen im Studienverlauf ab und verschlechterten sich dementsprechend bis zur letzten Beobachtung geringfügig (Tabelle 4-54).

Die Rücklaufquoten für die eingesetzten Instrumente lagen bei  $\geq 96\%$ .

Tabelle 4-54: Änderung des PedsQL™ von Baseline bis zur letzten Erhebung (ITT n=23, gesamtes Dosisintervall) [26,57]

Endpunkt	n	PedsQL™ Mittelwert (SD)			
		Baseline-Wert <sup>1</sup>	Wert in Woche 25	Wert zu Studienende oder Behandlungsende	Änderung bis zur letzten Erhebung <sup>2</sup>
Fremdbericht bei Kleinkindern (generischer Gesamtscore)	23	60,7 (12,80)	67,2 (12,72)	63,3 (15,23)	2,6 (12,16)
Körperliche Fähigkeiten	23	63,5 (22,15)	63,0 (24,00)	57,3 (23,55)	-6,1 (21,72)
Emotionale Fähigkeiten	23	70,2 (20,67) (n=22)	82,8 (12,66)	77,8 (16,08)	7,3 (26,22) (n=22)
Soziale Fähigkeiten	23	49,8 (15,85)	55,7 (15,83)	57,6 (15,29)	7,8 (16,22)
Schule	22	57,0 (19,64)	72,0 (16,97)	63,3 (23,66)	6,3 (29,82)
Psychosozialer Summenscore	23	59,0 (12,40)	69,7 (10,11)	67,0 (13,15)	7,9 (14,91)
„Family Impact Module“, Fremdbericht, Gesamtscore	23	61,4 (14,27)	66,5 (17,42)	65,1 (15,46)	3,7 (19,04)
Körperliche Funktion	23	63,4 (16,52)	65,0 (21,12)	65,0 (16,03)	1,6 (20,86)
Emotionale Fähigkeiten	23	54,3 (20,58)	61,7 (22,25) (n=22)	64,6 (18,70)	10,2 (22,99)
Soziale Fähigkeiten	23	63,3 (22,08)	69,3 (21,30) (n=22)	58,7 (21,21)	-4,6 (25,23)
Kognitive Fähigkeiten	23	62,4 (21,63)	69,5 (22,52) (n=22)	69,8 (22,94)	7,4 (32,82)
Kommunikative Fähigkeiten	23	64,9 (22,33)	68,6 (22,56) (n=25)	67,0 (20,18)	2,2 (28,00)
Summenscore (körperlich, emotional, sozial und kognitiv)	23	60,9 (14,33)	66,2 (17,63) (n=22)	64,8 (15,65)	4,0 (19,10)
<sup>1</sup> Baseline ist der letzte Wert vor der ersten Infusion (Studien-Baseline). <sup>2</sup> Woche 49 ab 300 mg Baseline Ein Score von 0 entspricht dem schlechtesten Wert und ein Score von 100 dem besten Wert. CSR 190-201: Tabelle 11.4.2.2.1, Tabelle 11.4.2.2.2, Tabelle 14.2.7.1, Tabelle 14.2.7.2					

### **Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext**

Da es kein validiertes Instrument zur Erfassung der Lebensqualität bei Patienten mit CLN2 gibt, wird erwartungsgemäß weder in Deutschland noch in anderen Ländern ein QoL-Instrument routinemäßig eingesetzt. Ähnliche Aspekte, wie z. B. Umgang der Familie mit der Erkrankung, werden im klinischen Alltag möglicherweise durch nicht zielgerichtete Fragen erfasst.

**CLN2-spezifische Lebensqualität – weitere Untersuchungen**

Tabelle 4-55: Operationalisierung des Endpunkts „CLN2-spezifische Lebensqualität“ – weitere Untersuchungen

Studie	Operationalisierung
190-201	<p>Der CLN2-spezifische Fragebogen zur Lebensqualität wurde von BioMarin entwickelt, um die speziellen Probleme von Angehörigen, die Kinder mit CLN2 betreuen, besser erfassen zu können. Hierzu wurden Eltern von Kindern mit CLN2 hinsichtlich ihrer Einschätzungen befragt. Der Fragebogen ist formell nicht vollständig für Patienten mit CLN2 validiert.</p> <p>Der CLN2-spezifische Fragebogen zur Lebensqualität umfasst 28 Items, die sechs Dimensionen abdecken:</p> <ul style="list-style-type: none"> <li>- Krampfanfälle (6 Items)</li> <li>- Ernährung / Keine Gastrostomie-Ernährungssonde (4 Items)</li> <li>- Schlaf (5 Items)</li> <li>- Verhalten (6 Items)</li> <li>- Tägliche Aktivitäten (4 Items)</li> </ul> <p>Jedes Item wurde auf einer 5-Punkt Likertskala von 0 (nie) bis 4 (fast immer) bewertet und jedes Item wurde in einen Score von 0-100 rücktransformiert: 0=100, 1=75, 2=50, 3=25, 4=0. Falls mehr als 50% der Items der Skala fehlten, sollten die Ergebnisse der Skala nicht analysiert werden.</p> <p>Aus der Summe der Zahl der beantworteten Items wurde der Mittelwert gebildet.</p> <p>Der Gesamtscore bildete die Summe aus der Zahl aller beantworteten Items aller Skalen.</p> <p>Daten zur Lebensqualität wurden vor der ersten Infusion (Baseline), zu Beginn der stabilen Dosisphase und danach alle 12 Wochen sowie zu Studienende oder bei vorzeitigem Abbruch erhoben. Die Fragebögen sollten möglichst vor den MRIs und den Probennahmen ausgefüllt werden. Baseline stellte die letzte Beobachtung vor der ersten 300 mg-Infusion dar; wenn diese nicht vorlag, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300 mg-Infusion eingegeben.</p> <p>Eine Imputation für fehlende Werte war nicht vorgesehen.</p> <p>Die berichteten Daten beziehen sich auf den finale Datenschnitt der Studie 190-201.</p>
190-202	<p>Der Endpunkt wurde wie oben beschrieben erhoben, allerdings fanden die Erhebungen während der Behandlung alle 24 Wochen und zu Behandlungsende statt. Die Ergebnisse des Endpunkts werden nicht berichtet, da lediglich eine Zwischenanalyse der fortlaufenden Studie vorliegt und zu diesem Zeitpunkt nicht alle Patienten in die Studie eingeschlossen worden waren.</p>
190-901 (natural history)	<p>Der Endpunkt wurde in der Studie 190-901 nicht erhoben.</p>

Es liegen bei den Patienten der historischen Kontrollgruppe keine Daten zur Lebensqualität vor und entsprechend auch keine Vergleiche zwischen der Studie 190-201 und der historischen Kontrolle. Hinsichtlich der Vollständigkeit der Daten, fehlten nur wenige Werte zu Baseline und bei der letzten Auswertung. Das Instrument erfasst direkte Beobachtungen und nicht subjektive Erfahrungen/Empfindungen der Kinder, wie dies bei sehr kleinen Kindern empfohlen wird [47]. Das Instrument ist bei CLN2-Patienten nicht validiert. Es ist allerdings wahrscheinlich, dass der Endpunkt für Eltern, die sich um Kinder mit CLN2 kümmern, von Bedeutung ist, da betroffene Eltern an der Entwicklung des Instruments aktiv beteiligt waren. Allerdings ist noch unklar, welches Ausmaß der Veränderung relevant ist.

Zwischen Baseline und Woche 49 der stabilen Dosisphase nahm der CLN2-spezifische QoL-Score um 8,1 (14,33) Punkte zu, was einer Verbesserung um etwa 10,9% entspricht. Die Behandlung mit Cerliponase alfa hatte keinen negativen Effekt auf irgendeine der Einzeldomänen. Mit Ausnahme der Domäne Alltagsaktivitäten (vernachlässigbarer Abfall), verbesserten sich alle Domänen und insbesondere die Krampffneigung der Patienten (Tabelle 4-56).

Tabelle 4-56: Änderung der gesundheitsbezogenen Lebensqualität von Baseline bis zur letzten Erhebung gemessen mittels eines CLN2-spezifischen Instruments (gesamtes Dosisintervall) [26,57]

Endpunkt	n	CLN2QoL™ (Punkte) Mittelwert (SD)			
		Baseline-Wert <sup>1</sup>	Wert in Woche 25	Wert zu Studienende oder Behandlungsende	Änderung bis zur letzten Erhebung
CLN2-spezifische Lebensqualität, Gesamt score	22	74,2 (13,82)	82,4 (8,59)	81,9 (11,10) (n=23)	8,1 (14,33)
Krampfanfälle	22	64,2 (18,17)	79,4 (12,10)	82,8 (13,19) (n=23)	18,2 (23,02)
Ernährung (ohne Gastrostomiesonde)	22	75,6 (22,57)	78,6 (17,29) (n=21)	83,2 (16,75)	9,8 (22,15) (n=21)
Schlaf	22	79,8 (25,89)	91,4 (12,74)	83,0 (16,97) (n=23)	4,1 (19,74)
Verhalten	22	73,7 (13,57)	78,0 (14,78)	80,2 (17,25) (n=23)	7,2 (18,01)
Alltagsaktivitäten	22	81,7 (15,35)	85,8 (14,71)	80,7 (19,21) (n=23)	-0,8 (18,10)

<sup>1</sup> Baseline ist der letzte Wert vor der ersten Infusion.  
Ein Score von 0 entspricht dem schlechtesten Wert und ein Score von 100 dem besten Wert.  
CSR 190-201: Tabelle 14.2.7.3

### **Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext**

Da es kein validiertes Instrument zur Erfassung der Lebensqualität bei Patienten mit CLN2 gibt, wird erwartungsgemäß weder in Deutschland noch in anderen Ländern ein QoL-Instrument routinemäßig eingesetzt. Ähnliche Aspekte, wie z. B. die Belastung der Familie durch die Erkrankung, werden im klinischen Alltag möglicherweise durch nicht zielgerichtete Fragen erfasst.

#### **4.4.2.3.3 Verträglichkeit und Sicherheit (UEs) – weitere Untersuchungen**

Tabelle 4-57: Operationalisierung des Endpunkts Verträglichkeit und Sicherheit

Studie	Operationalisierung
190-201	Wenn nicht anders angegeben beziehen sich die Angaben auf Amendment 5 des Studienprotokolls [37] und den SAP [51].



Studie	Operationalisierung
	<p><b>Definition von UEs</b></p> <p>Es wurden behandlungsbezogene UEs dokumentiert. Diese schlossen jedes neu aufgetretene Ereignis, jedes Ereignis mit erhöhter Frequenz und jedes Ereignis, dessen Schwere nach der ICV-Implantation zugenommen hatte, ein, unabhängig davon, ob der Patient Cerliponase alfa erhalten hatte oder nicht. Die UEs wurden getrennt nach dem Zeitpunkt ihres Auftretens dargestellt (nach der ICV-Implantation aber vor der ersten Infusion oder nach der ersten Infusion).</p> <p>Vorzugsweise sollte eine Diagnose statt einzelner, mit dieser Diagnose in Beziehung stehender UEs berichtet werden.</p> <p>Der Prüfarzt bewertete den Schweregrad jedes Ereignisses nach NCI CTCAE, Version 4. UEs ohne CTCAE-Term wurden entsprechen der allgemeinen CTCAE-Kriterien (Version 4) klassifiziert (Grad 1: Mild, Grad 2: Moderat, Grad 3: Schwerwiegend oder medizinisch signifikant aber nicht unmittelbar lebensbedrohlich, Grad 4: Lebensbedrohliche Konsequenzen, dringende Intervention erforderlich, Grad 5: Tod durch UE). Falls ein UE mehr als einmal auftrat wurde das jeweils schwerste Ereignis eingegeben.</p> <p>Unter folgenden Bedingungen wurde ein UE als behandlungsbedingt eingestuft: Der Zeitabstand zwischen dem Auftreten des Ereignisses und der Verabreichung des Prüfpräparates erscheint plausibel und es besteht ein begründeter Verdacht, dass das Ereignis mit dem Prüfpräparat in Zusammenhang steht. Der Prüfarzt sollte bewerten, ob das UE mit dem implantierten ICV-Zugang in Beziehung stand oder nicht (Letzteres mit Amendement 3 des Studienprotokolls eingeführt; 07.07.2014).</p> <p>Alle UEs wurden nach MedDRA Version 18.1 kodiert.</p> <p><b>Definition von SUEs</b></p> <p>SUEs waren wie folgt definiert.</p> <ul style="list-style-type: none"> <li>- Es ist fatal.</li> <li>- Es ist lebensbedrohlich.</li> <li>- Anmerkung: Lebensbedrohlich bedeutet in diesem Zusammenhang, dass der Patient aufgrund des Ereignisses zu versterben drohte, jedoch nicht Ereignisse, die hypothetisch zum Tode hätten führen können, wären sie schwerer gewesen.</li> <li>- Es erfordert oder verlängert einen stationären Aufenthalt.</li> <li>- Es führt zu anhaltender oder bedeutender Einschränkung oder Behinderung</li> <li>- Es handelt sich um eine angeborene Anomalie oder einen Geburtsfehler eines Kindes oder Fötus nach Exposition mit dem Prüfpräparat vor der Konzeption oder während der Schwangerschaft.</li> <li>- Es handelt sich um ein medizinisch bedeutsames Ereignis oder eine Reaktion.</li> </ul> <p>CTCAE Grad 4 oder 5 UEs waren <i>per definitionem</i> SUEs.</p> <p><b>Vordefinierte UEs besonderen Interesses</b> (“adverse events of special interest” AEOSI):</p> <ul style="list-style-type: none"> <li>- Überempfindlichkeitsreaktionen: Standardisierte MedDRA-Abfragen („Standardised MedDRA query“, SMQ) “Überempfindlichkeit” und “Anaphylaktische Reaktionen”. Der Algorithmus “Anaphylaktische Reaktionen” erforderte mindestens eines der folgenden Ereignisse: Ein “A” Ereignis (enger Umfang), oder ein “B und C” Ereignis (B: obere Atemwege / respiratorisch; C: Angioödem / Urtikaria / Pruritus / Hautrötung), oder ein “B und D” Ereignis (D: kardiovaskulär / Hypotonie), oder ein “C und D” Ereignis, welches innerhalb von 24 Stunden nach Beginn der Infusion auftrat.</li> <li>- Reaktionen, die in zeitlichem Zusammenhang mit der Infusion auftraten („temporally related event“, TRE): Jedes UE, einschließlich Anaphylaxie oder andere Überempfindlichkeits- oder allergische Reaktionen, das innerhalb von 24 Stunden nach Beginn oder Neubeginn der Cerliponase alfa-Infusion auftrat, unabhängig von der Kausalitätsbewertung des Prüfarztes.</li> <li>- Status epilepticus (MedDRA PTs: Petit mal Epilepsie, Status epilepticus)</li> </ul>

Studie	Operationalisierung
	<ul style="list-style-type: none"> <li>- Hydrozephalus (kommunizierender und nicht kommunizierender): MedDRA PT Hydrozephalus, kongenitaler Hydrozephalus</li> <li>- Meningitis: Entsprechende MedDRA PTs waren im SAP definiert.</li> <li>- Unerwarteter, rascher, nicht anders erklärbarer Abfall des CLN2-Scores: Identifiziert durch klinische Bewertung</li> </ul> <p><b>Berichtszeitraum</b></p> <p>UEs und SUEs waren wie folgt zu berichten:</p> <ul style="list-style-type: none"> <li>- Nach der Einverständniserklärung nur SUEs, die mit studienbedingten Massnahmen in Zusammenhang standen</li> <li>- Alle UEs/SUEs nach Implantation des ICV-Zugangs (behandlungsbezogene UEs).</li> </ul> <p>Falls der Zeitpunkt des Auftretens eines UEs unbekannt war, wurde das UE als behandlungsbezogen eingestuft. Die Nachbeobachtung betrug sechs Monate nach der letzten Infusion oder einem vorzeitigem Abbruch. Die Nachbeobachtung entfiel, wenn der Patient innerhalb von vier Wochen nach der letzten Infusion in die Folgestudie 190-202 eingeschlossen wurde.</p> <p>Mit dem ICV-Zugang in Zusammenhang stehende SUEs waren der Pharmakovigilanz innerhalb von 24 Stunden zu berichten. Dies betraf alle zur Infusion benötigten Komponenten. Die Patienten wurden angehalten, jedes nachfolgende SUE, welches aus Sicht des weiterbehandelnden Arztes mit der vorherigen Studienmedikation in Zusammenhang stand, zu melden. Die Prüfarzte sollten jeden späteren Todesfall und jedes SUE mitteilen, welches ihrer Ansicht nach mit der vorherigen Behandlung in Zusammenhang stehen könnte. Der Sponsor sollte ebenfalls kontaktiert werden, wenn bei einem Patienten nach Ende der Studie Krebs aufgetreten war oder eine angeborene Anomalie bei einem nach Abschluss der Studie gezeugten Kind.</p> <p>Ereignisse, die dringende Sicherheitsmaßnahmen erforderten (Richtlinie 2001/20/EC), waren vom Zeitpunkt der Einverständniserklärung bis zur letzten Visite zu dokumentieren.</p> <p>UEs wurden während der Studie kontinuierlich überwacht.</p> <p><b>Stopregeln</b></p> <p>Nicht akzeptable, behandlungsbedingte Toxizität war prospektiv wie folgt definiert: Behandlungsbedingte, Grad <math>\geq 3</math> Überempfindlichkeits- oder allergische Reaktionen, die auf medizinische Interventionen nicht ansprachen oder trotz geeigneter Vorsichtsmaßnahmen (Vorbehandlung oder Änderung der Infusionsrate) während nachfolgender Infusion in ähnlicher oder höherer Intensität wieder auftraten.</p> <p>Unter folgenden Bedingungen konnte die Rekrutierung nach Absprache mit dem DMC ausgesetzt werden: Nach Einschluss von <math>\leq 10</math> Patienten entwickeln zwei oder mehr Patienten eine nicht akzeptable Toxizität; nach Einschluss von <math>&gt;10</math> aber <math>&lt;20</math> Patienten tritt bei drei oder mehr Patienten eine nicht akzeptable Toxizität auf.</p> <p>Für alle anderen Ereignisse galt Folgendes:</p> <ul style="list-style-type: none"> <li>- Falls ein Patient ein CTCAE Grad <math>\geq 3</math> UE entwickelte, welches aus Sicht des Prüfarztes nicht durch die Grunderkrankung bedingt war, konnte je nach Art und Dauer des Ereignisses ein Ausschluss des Patienten erwogen werden.</li> <li>- Falls bei zwei Patienten dasselbe CTCAE Grad 3 UE oder bei einem Patienten ein Grad 4 UE auftrat und ein Zusammenhang mit der Grunderkrankung nicht vermutet wurde, sollte das Risiko-Nutzen-Verhältnis der Studie erneut überprüft werden.</li> </ul> <p><b>Besondere Überwachungsmaßnahmen</b></p> <p>Aus Sicherheitsgründen wurde Cerliponase alfa im Rahmen der Studie nur stationär verabreicht. Die Routineüberwachung der Studie machte folgende Sicherheitsmaßnahme erforderlich (Amendment 3 des Studienprotokolls, 03.07.2014): Nach der Implantation, der ersten Dosis und Dosisescalation wurden die Patienten über 48 stationär überwacht, davon 24 Stunden in der pädiatrischen Intensivstation. Innerhalb von 24 Stunden nach der Entlassung wurden die Patienten telefonisch kontaktiert. Mit Amendment 5 des Studienprotokolls (24.09.2014) durften Patienten während der stabilen Dosisphase nach 24 Stunden entlassen werden, vorausgesetzt sie waren klinisch stabil und es bestanden keine sonstigen Sicherheitsbedenken.</p>

Studie	Operationalisierung
	<p>Aufgrund des Risikos von Überempfindlichkeitsreaktionen stand eine Notfallausrüstung während der Infusionen bereit und entsprechend ausgebildetes Personal war anwesend.</p> <p>Der Prüfarzt hatte BioMarin so schnell wie möglich mindestens innerhalb von 24 Stunden über schwere UEs (Grad <math>\geq 3</math>) oder SUEs zu informieren, falls</p> <ul style="list-style-type: none"> <li>- die Infusion unterbrochen, abgebrochen oder verändert worden war (nicht bedingt durch einen Verschluss des Zugangs).</li> <li>- Flüssigkeiten einschließlich Steroide und Anti-Histaminika intravenös verabreicht worden waren.</li> <li>- der Patient mit Sauerstoff behandelt worden war.</li> </ul> <p>BioMarin war über jedes der UEs besonderen Interesses innerhalb von 24 Stunden zu unterrichten.</p> <p>Mit Amendement 2 des Studienprotokolls (26.07.2013, in Kraft vor Einschluss des ersten Patienten) wurde darauf hingewiesen, dass die Patienten während der Infusionen regelmäßig durch ausgebildetes Personal zu überwachen sind (UEs und epileptische Anfälle) und die Infusionen im Ermessen des Prüfarztes unterbrochen werden können. Vor jeder Infusion war die Funktion des ICV-Zugangs zu überprüfen (Durchgängigkeit, Lokalisation und Hautzustand).</p> <p>Während der Studie wurde besonders auf mögliche Infektionen (erhöhte Temperatur, Husten, Ausschlag, Kopfschmerzen, Änderung des Gemütszustands, Schwellung oder Drainage an der Einstichstelle) und Anzeichen für ein Leck oder Versagen des ICV-Zugangs geachtet.</p> <p><b>Patientenrelevante UEs</b></p> <p>Die folgenden patientenrelevanten UEs werden berichtet:</p> <ul style="list-style-type: none"> <li>- Gesamtzahl der UEs</li> <li>- Gesamtzahl der UEs pro SOK</li> <li>- Häufige UEs</li> <li>- SUEs und Todesfälle</li> <li>- CTCAE Grad <math>\geq 3</math> UEs</li> <li>- Studien- oder Behandlungsabbruch aufgrund eines UEs</li> <li>- UEs besonderen Interesses (AEOSI): <ul style="list-style-type: none"> <li>• Überempfindlichkeits- und anaphylaktische Reaktionen (MedDRA SMQs)</li> <li>• Infusionsassoziierte UEs (TRE)</li> <li>• Erkrankungen des Nervensystems, MedDRA SMQ "Konvulsionen", Status epilepticus, Hydrozephalus und Meningitis</li> <li>• Unerwarteter und rascher, nicht anders erklärbarer Abfall des CLN2-Scores (ML)</li> </ul> </li> <li>- UEs, die mit dem ICV-Zugang in Zusammenhang standen.</li> </ul> <p>Die Sicherheitsdaten wurden deskriptiv berichtet. Dies ist angemessen, da der Studie keine sicherheitsrelevanten Hypothesen zugrunde lagen [54], auch wenn die Sicherheit und Verträglichkeit eines der primären Ziele der Studie war.</p>

Studie	Operationalisierung
190-202	<p>Die Sicherheit und Verträglichkeit wurde wie oben beschrieben erhoben, abgesehen von folgenden Änderungen:</p> <p>Alle UEs, die mit dem ICV-Zugang in Zusammenhang standen, waren als UEs besonderen Interesses definiert. Die übrigen UEs besonderen Interesses entsprachen denjenigen der Studie 190-201.</p> <p>Alle UEs und SUEs, die nach der ersten Dosis in der Studie 190-202 aufgetreten waren, wurden dokumentiert. SUEs waren vom Zeitpunkt der Einwilligungserklärung bis sechs Monate nach der letzten Dosis oder nach Studienabbruch zu berichten, je nachdem welcher Zeitraum länger war. Die Nachbeobachtung entfiel, wenn der Patient an einer anderen BioMarin-Studie oder einem Register teilnahm. Fortbestehende, behandlungsbedingte UEs und alle andauernden SUEs, Schwangerschaften und mit dem ICV-Zugang in Zusammenhang stehende UEs wurden bis zum vollständigen Abklingen beobachtet. Alle anderen UEs einschließlich solcher UEs, die bereits in der Studie 190-202 aufgetreten waren und sich nach der ersten Dosis in der Folgestudie verschlechtert hatten, wurden in der neuen Studie erfasst.</p> <p>UEs, die nach Einschätzung des Prüfarztes mit dem ICV-Zugang in Zusammenhang standen, waren innerhalb von 24 Stunden an BioMarin (Pharmakovigilanz) zu melden.</p> <p>Im Protokoll waren Bedingungen festgelegt, unter welchen eine klinische Laborabweichung als UE zu dokumentieren war.</p> <p><b>Abbruchregeln</b></p> <p>Ein Patient, bei dem eine nicht akzeptable (CTCAE-Grad <math>\geq 3</math>), nach Ansicht des Prüfarztes behandlungsbedingte Überempfindlichkeits- oder allergische Reaktion auftrat, konnte unter folgenden Bedingungen nach Rücksprache ausgeschlossen werden:</p> <ul style="list-style-type: none"> <li>- Keine Besserung nach Einsatz medizinischer Maßnahmen wie z.B. Unterbrechung der Infusion, Reduktion der Infusionsrate, i.v. Anti-Histaminika, Sauerstoff, i.v. Flüssigkeiten, oder Steroide oder</li> <li>- Erneute ähnlich schwere oder schwerere Reaktion auf eine nachfolgende Infusion trotz angemessener Vorbehandlung und Änderung der Infusionsrate.</li> <li>- Patienten, die innerhalb eines Jahres drei oder mehr Punkte auf der ML-Skala verloren hatten, konnten im Ermessen des Prüfarztes aus der Studie ausgeschlossen werden.</li> <li>- Patienten mit einem ML-Score von Null bei zwei aufeinander folgenden Visiten wurden aus der Studie ausgeschlossen.</li> </ul> <p>Zur Sicherheit wird der aktuellste Datenschnitt (Juni 2016) der Studie berichtet.</p>
190-901 (natürlicher Verlauf)	Sicherheitsdaten wurden in der Studie nicht erfasst.

Die Studie entsprach den Vorgaben der GCP-Verordnung und die Endpunkte zur Sicherheit und Verträglichkeit wurden nach aktuellen, international gültigen Standards erhoben und dargestellt. Die UEs wurden nach MedDRA Version 18.0 (190-201) oder 19.1 (letzte Auswertung der Studie 190-202). Zur Übersetzung ins Deutsche wurde die Datenbank EUCTC verwendet, um eine korrekte Kodierung der UEs zu gewährleisten. UEs von speziellem Interesse waren prospektiv definiert. Die Methoden zur Bewertung und Berichterstattung der UEs stimmen mit den Standards des TREND-Statements überein.

Alle UEs, die nach Implantation des ICV-Ports auftraten, wurden erfasst. Alle Patienten wurden in die Analyse zur Sicherheit eingeschlossen, darunter auch ein Patient, der die Studie nach nur

einer Infusion aus Gründen abgebrochen hatte, die nicht mit der Sicherheit in Zusammenhang standen.

Die Prüfarzte waren angehalten, speziell auf solche UEs zu achten, die in einem engen, zeitlichen Zusammenhang mit den Cerliponase alfa-Infusionen auftraten. Die Infusionen wurden ausschließlich im Krankenhaus in Anwesenheit entsprechend geschulten Personals verabreicht. Die Patienten wurden während und nach den Infusionen besonders hinsichtlich epileptischer Krampfanfälle und Überempfindlichkeitsreaktionen überwacht. Mit Amendement 4 des Studienprotokolls (08.08.2014) wurde den Patienten und deren Betreuern nach Implantation des ICV-Ports ein Dokument ausgehändigt, welches Anzeichen oder Symptome, die auf mögliche Komplikationen oder ein Versagen des Ports hindeuten, detailliert beschreibt. Nach einer Erweiterung des CONSORT-Statements kann die Erwartung eines UEs die Inzidenz der gemeldeten UEs beeinflussen [19]. In der Studie 190-201 wurde Cerliponase alfa als ein Wirkstoff biologischen Ursprungs erstmalig bei Menschen eingesetzt. Daher war eine engmaschige, stationäre Überwachung der Patienten aus Gründen der Sicherheit erforderlich. Mit Amendement 3 des Studienprotokolls vom 07.07.2014 wurden die Prüfarzte gebeten, einen möglichen Zusammenhang zwischen UEs und dem ICV-Port zu bewerten.

Die Sicherheit wurde wie im Protokoll prospektiv festgelegt analysiert, abgesehen von geringfügigen Änderungen. Neben Status epilepticus (MedDRA PT Petit mal Epilepsie und Status epilepticus), welcher im SAP als UE von besonderem Interesse definiert war, wurde die breite MedDRA SMQ „Konvulsionen“ eingesetzt. UEs in Zusammenhang mit dem ICV-Port galten protokollgemäß nicht als UEs von besonderem Interesse, allerdings wurden derartige UEs wie z. B. Infektionen des Ports erwartet und die Sicherheit und Funktion des ICV-Ports waren regelmäßig zu kontrollieren. Dies ist nicht als selektive Berichterstattung zu bewerten, sondern dient ausschließlich dazu, das Sicherheitsprofil von Cerliponase sowie des erforderlichen ICV-Ports weiter zu charakterisieren. Mit der nicht geplanten Zwischenanalyse der Studie 190-202 (Datenschnitt Juni 2016) wurde eine Forderung der FDA umgesetzt. Es liegen keine Hinweise vor, dass Endpunkte zur Sicherheit oder Verträglichkeit ergebnisgesteuert berichtet wurden, und auch keine sonstigen Verzerrungsquellen, abgesehen von dem grundsätzlich hohen Verzerrungspotential nicht kontrollierter Studien.

Nach Studienbeginn gab es einige, sicherheitsrelevante Änderungen. Dies ist allerdings nicht ungewöhnlich, da in der Studie 190-201 erstmalig ein neuer Wirkstoff biologischen Ursprungs beim Menschen eingesetzt wurde, der zudem noch ICV zu verabreichen ist. Fast alle sicherheitsrelevanten Änderungen erfolgten auf Empfehlung der Zulassungsbehörden oder der lokalen Ethikkommissionen.

Ab Amendement 2 (vom 26. Juli 2013) durften die Patienten im Ermessen des Prüfarztes mit geeigneten, altersgerechten Sedativa (z. B. vor bildgebenden Verfahren des Gehirns) vorbehandelt werden. Der erste Patient wurde eingeschlossen, als Amendement 2 gültig war. Einige Patienten wurden vor einzelnen Cerliponase alfa-Infusionen nicht protokollgemäß vorbehandelt (zur Vermeidung von Überempfindlichkeitsreaktionen), was das Risiko, ein UE zu entwickeln, erhöht haben könnte.

Insgesamt wurden die Sicherheit und Verträglichkeit von Cerliponase und dem ICV-Port entsprechend der aktuellen Standards erhoben und ausgewertet.

Zur Inzidenz und Schwere der UEs werden die Ergebnisse aus dem Abschlussbericht der Studie 190-201 und der jüngsten Zwischenanalyse der Studie 190-202 dargestellt (Datenschnitt Juni 2016, 72 Wochen Nachbeobachtung). Es erschien nicht sinnvoll, die Ergebnisse einer früheren Zwischenanalyse der Studie 190-202 darzustellen (Datenschnitt: 15.10.2015; 5/23 Patienten der Studie 190-201 waren zu diesem Zeitpunkt noch nicht in Studie eingeschlossen worden), obwohl diese den Behörden im Rahmen der Zulassung übermittelt wurde.

Die Patienten wurden im Rahmen der Studie 190-201 im Mittel (SD) über 49,5 (13,23) Wochen mit Cerliponase alfa behandelt und über 96,4 (29,77) Wochen in der jüngsten Analyse der Studie 190-202 (Tabelle 14.3.6.1.1).

Die MedDRA SOK „Produktprobleme“ wurde erst im Jahr 2016 eingeführt und wurde daher im Abschlussbericht der Studie 190-201 nicht zur Kodierung eingesetzt, allerdings in der jüngsten Analyse der Studie 190-202. Daher wurden einige der UEs (PTs) in den Analysen unterschiedlich kodiert, z. B. einige der UEs im Zusammenhang mit dem ICV-Port unter der SOK „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ im Studienbericht 190-201 und unter der SOK „Produktprobleme“ in der jüngsten Analyse der Studie 190-202.

Tabelle 4-58 gibt einen Überblick der UEs, die während der Dosisescalation, stabilen Dosisphase und in der Gesamtpopulation aufgetreten sind. Da nur wenige Patienten an der Dosisescalation teilgenommen hatten und diese Daten keine zuverlässigen Schlussfolgerungen zulassen, werden die Ergebnisse zur Sicherheit und Verträglichkeit nachfolgend nicht getrennt für die Dosiskohorten und die stabile Dosisphase dargestellt. Hinsichtlich dieser Daten wird auf den Studienbericht 190-201 verwiesen [26].

Tabelle 4-58: Überblick der in der Studie 190-201/202 aufgetretenen, unerwünschten Ereignisse in den Kohorten 1-3, während der stabilen Dosierung und in der Gesamtpopulation [26]

Kategorie	Studie 190-201 (Finaler Datenschnitt) n=24				
	Kohorte 1 (n=3)	Kohorte 2 (n=3)	Kohorte 3 (n=4)	Stabile Dosierung (n=14)	Gesamt (n=24)
Mindestens 1 UE	3 (100%)	3 (100%)	4 (100%)	14 (100%)	24 (100%)
Mindestens ein mit der Behandlung in Zusammenhang stehendes UE	3 (100%)	3 (100%)	4 (100%)	13 (93%)	23 (96%)
Mindestens 1 SUE	1 (33%)	3 (100%)	3 (75%)	9 (64%)	16 (67%)
Mindestens ein mit der Behandlung in Zusammenhang stehendes SUE	1 (33%)	3 (100%)	1 (25%)	2 (14%)	7 (29%)
Mindestens 1 UE, welches zum Studienabbruch führte	0	0	0	0	0
Mindestens 1 UE, welches zum endgültigen Studienabbruch führte	0	0	0	0	0
Todesfälle	0	0	0	0	0
CSR 190-201: Tabelle 12.2.1.1					

### Gesamtzahl der UEs

Tabelle 4-59: Anzahl der Patienten, die mindestens ein unerwünschtes Ereignis entwickelten („Sicherheits“-Population) [26,34]

Kategorie	Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt	Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016	Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016
Patienten mit mindestens einem UE	24 (100%)	24 (100%)	23 (100%)
Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1.			
CSR 190-201: Tabelle 12.2.1.1			
FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.1			

Wie bei einer derartig schweren Erkrankung und bei einem vulnerablen, pädiatrischen Patientenkollektiv zu erwarten, trat bei jedem der 24 Patienten (100%) während der Studie mindestens ein UE auf. Die Anzahl der Patienten, die mindestens ein UE entwickelten, blieb

im Studienverlauf weitgehend konstant, unabhängig von der Dauer der Behandlung: >0-24 Wochen (n=24/24, 100%), >24-48 Wochen (n=23/23, 100%), >48-72 Wochen (n=22/23, 96%), >72-96 Wochen (n=20/23, 87%) und >96 Wochen (n=10/11, 91%) (FDA Response Datenschnitt: June 2016: Tabelle 14.3.1.2.9).

### ***Inzidenz von UEs getrennt nach SOK***

Tabelle 4-60: Inzidenz unerwünschter Ereignisse stratifiziert nach Systemorganklassen („Sicherheits“-Population) [26,34]

<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Erkrankungen des Nervensystems	24 (100%)	24 (100%)	17 (74%)
Infektionen und parasitäre Erkrankungen	23 (96%)	24 (100%)	20 (87%)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	16 (67%)	19 (79%)	14 (61%)
Erkrankungen des Gastrointestinaltrakts	19 (79%)	19 (79%)	9 (39%)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	17 (71%)	19 (79%)	3 (13%)
Psychiatrische Erkrankungen	12 (50%)	12 (50%)	3 (13%)
Erkrankungen der Atemwege, des Brustraums und Mediastinums	6 (25%)	10 (42%)	2 (9%)
Erkrankungen des Immunsystems	8 (33%)	10 (42%)	0 (0%)
Stoffwechsel- und Ernährungsstörungen	5 (21%)	9 (38%)	5 (22%)
Produktprobleme	0 (0%)	7 (29%)	5 (22%)
Erkrankungen der Haut und des Unterhautzellgewebes	5 (21%)	6 (25%)	3 (13%)
Untersuchungen	3 (13%)	6 (25%)	5 (22%)
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen	1 (4%)	5 (21%)	4 (17%)
Augenerkrankungen	3 (13%)	4 (17%)	1 (4%)
Erkrankungen des Blutes und des Lymphsystems	3 (13%)	4 (17%)	0 (0%)
Gefäßerkrankungen	2 (8%)	4 (17%)	1 (4%)
Herzkrankungen	4 (17%)	4 (17%)	0 (0%)
Chirurgische und medizinische Eingriffe	1 (4%)	2 (8%)	0 (0%)



<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Endokrine Erkrankungen	0 (0%)	1 (4%)	1 (4%)
Erkrankungen der Niere und Harnwege	0 (0%)	1 (4%)	1 (4%)
Erkrankungen der Geschlechtsorgane und der Brustdrüse	1 (4%)	1 (4%)	0 (0%)
<p>Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1.</p> <p>Die deutschen Übersetzungen der MedDRA SOKs und PTs wurden in der Datenbank EUTCT recherchiert.</p> <p>CSR 190-201: Tabelle 14.3.1.2.1</p> <p>FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.2.1, Tabelle 14.3.1.13.1</p>			

Folgende SOKs wurden am häufigsten mit einer Inzidenz von  $\geq 10\%$  berichtet (Abschlussbericht 190-201 bzw. Analyse 190-202 mit Datenschnitt im Juni 2016): „Erkrankungen des Nervensystems“ (100% / 100%), „Infektionen und parasitäre Erkrankungen“ (96% / 100%), „Erkrankungen des Gastrointestinaltrakts“ (79% / 79%) „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ (67% / 79%), „Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen“ (71% / 79%), „Psychiatrische Erkrankungen“ (50% / 50%), „Erkrankungen des Immunsystems“ (33% / 42%), „Erkrankungen der Atemwege, des Brustraums und Mediastinums“ (25% / 42%), „Stoffwechsel- und Ernährungsstörungen“ (21% / 38%), „Produktprobleme“ (0% / 29%), „Erkrankungen der Haut und des Unterhautgewebes“ (21% / 25%), „Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen“ (4% / 21%), „Augenerkrankungen“ (13% / 17%), und „Erkrankungen des Blutes und des Lymphsystems“ (8% / 17%; Tabelle 4-60).

Wie der Vergleich der Studie 190-201 mit der aktuellen Analyse der Studie 190-202 zeigt, bleibt die Inzidenz der UEs pro SOK im Studienverlauf weitgehend stabil.

**Inzidenz sehr häufiger UEs**Tabelle 4-61: Inzidenz sehr häufiger unerwünschter Ereignisse ( $\geq 10\%$  der Patienten) mit Angabe der Inzidenz schwerer UEs (CTCAE Grad  $\geq 3$ ; „Sicherheits“-Population) [26,34]

UE (MedDRA PT)	Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt	Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016	Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016
Fieber <sup>2</sup>	13 (54%)	17 (71%)	12 (52%)
Krampfanfall Grad 3	14 (58%)	14 (58%) 1 (4%)	7 (30%)
Erbrechen <sup>2</sup>	11 (46%)	12 (50%)	5 (22%)
Infektion der oberen Atemwege	10 (42%)	12 (50%)	6 (26%)
Epilepsie	11 (46%)	11 (46%)	5 (22%)
Generalisierte tonisch-klonische Anfälle	n.b. oder <10%	11 (46%)	9 (39%)
Rhinitis	5 (21%)	10 (42%)	6 (26%)
Nasopharyngitis	7 (29%)	9 (38%)	5 (22%)
Überempfindlichkeit <sup>2</sup> Grad 3	8 (33%) 3 (13%)	9 (38%) 3 (13%)	n.b. oder <10%
Obstipation	7 (29%)	8 (33%)	n.b. oder <10%
Virusinfektion	5 (21%)	8 (33%)	n.b. oder <10%
Gangstörung	7 (29%)	7 (29%)	n.b. oder <10%
Gastroenteritis	5 (21%)	7 (29%)	4 (17%)
Husten	5 (21%)	7 (29%)	n.b. oder <10%
Myoklonus	7 (29%)	7 (29%)	n.b. oder <10%
Petit mal epilepsy	n.b. oder <10%	7 (29%)	3 (13%)
Sturz	6 (25%)	7 (29%)	n.b. oder <10%
Tremor	4 (17%)	7 (29%)	n.b. oder <10%
Diarrhö	4 (17%)	6 (25%)	3 (13%)
Pharyngitis Grad 3	6 (25%) 1 (4%)	6 (25%) 1 (4%)	n.b. oder <10%
Virusinfektion der oberen Atemwege	n.b. oder <10%	6 (25%)	6 (26%)
Babinski Reflex	4 (17%)	5 (21%)	n.b. oder <10%
Dysphagie Grad 3	4 (17%)	5 (21%) 1 (4%)	n.b. oder <10%
Dystonie	4 (17%)	5 (21%)	n.b. oder <10%

UE (MedDRA PT)	Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt	Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016	Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016
Problem mit einer Nadel	n.b. oder <10%	5 (21%) <sup>1</sup>	n.b. oder <10%
Tonsillitis		5 (21%)	4 (17%)
Anomales Verhalten	3 (13%)	4 (17%)	n.b. oder <10%
Entwicklungsverzögerung	3 (13%)	4 (17%)	n.b. oder <10%
Infektion der Atemwege	n.b. oder <10%	4 (17%)	n.b. oder <10%
Kopfschmerz <sup>2</sup>	3 (13%)	4 (17%)	n.b. oder <10%
Konjunktivitis	<10%	4 (17%)	n.b. oder <10%
Reizbarkeit <sup>2</sup>	4 (17%)	4 (17%)	n.b. oder <10%
Schlaflosigkeit	4 (17%)	4 (17%)	n.b. oder <10%
Agitiertheit	n.b. oder <10%	3 (13%)	n.b. oder <10%
Athetose (nicht in EUTCT identifiziert)	n.b. oder <10%	3 (13%)	n.b. oder <10%
Harnwegsinfektion	n.b. oder <10%	3 (13%)	n.b. oder <10%
Kopfverletzung	n.b. oder <10%	3 (13%)	n.b. oder <10%
Muskelspasmen	n.b. oder <10%	3 (13%)	
Pleozytose (nicht in EUTCT identifiziert)	3 (13%)	3 (13%)	n.b. oder <10%
Rhinovirus-Infektion	n.b. oder <10%	3 (13%)	n.b. oder <10%
Schlafstörung	3 (13%)	3 (13%)	n.b. oder <10%
Verminderter Appetit	n.b. oder <10%	3 (13%)	n.b. oder <10%
Zahnkaries	n.b. oder <10%	3 (13%)	n.b. oder <10%
<p><sup>1</sup> Das UE wurde im CSR 190-201 unter der SOK „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ kodiert und zum Datenschnitt Juni 2016 unter der SOK „Produktprobleme“.</p> <p><sup>2</sup> Diese UEs sind in der Fachinformation als sehr häufige UEs von Cerliponase alfa aufgeführt.</p> <p>n.b. nicht berichtet</p> <p>Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1.</p> <p>CSR 190-201: Tabelle 12.2.3.1.1, Tabelle 14.3.1.2.2, Tabelle 14.3.1.2.3</p> <p>FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.2.3, Tabelle 14.3.1.13.3</p>			

Folgende UEs wurden am häufigsten berichtet ( $\geq 40\%$  bei mindestens einer der Analysen; Abschlussbericht 190-201 bzw. Analyse 190-202 mit Datenschnitt im Juni 2016): Fieber (54% / 71%), Krampfanfall (beide 58%), Erbrechen (46% / 50%), Infektion der oberen Atemwege (42% / 50%), Generalisierte tonisch-klonische Anfälle (8% / 46%), Epilepsie (beide 46%) und

Rhinitis (21% / 42%). Wie Tabelle 4-61 zeigt, traten auch eine Reihe anderer UEs sehr häufig auf.

Bei vier Patienten (17%) wurde „Entwicklungsverzögerung“ als UE berichtet, davon stand allerdings, nach Einschätzung der Prüfarzte, keines mit der Behandlung in Zusammenhang (FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.3.2).

### ***Inzidenz schwerwiegender UEs und Todesfälle***

Tabelle 4-62: Inzidenz schwerwiegender UEs und Todesfälle („Sicherheits“-Population)  
[26,27,34]

<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Mindestens ein SUE	16 (67%) 32	19 (79%) 51	6 (26%) 7
SUEs, die bei mindestens zwei Patienten oder mehrmals auftraten			
Überempfindlichkeit	6 (25%)	7 (29%) 9	≤2 oder n.b.
Infektion der oberen Atemwege	≤2 oder n.b.	4 (17%) 4	3 (13%) 3
Gastroenteritis	≤2 oder n.b.	3 (13%) 3	≤2 oder n.b.
Fieber	2 (8%)	2 (8%) 2	≤2 oder n.b.
Pharyngitis bakteriell	2 (8%)	2 (8%) 3	≤2 oder n.b.
Epilepsie	2 (8%)	2 (8%) 2	≤2 oder n.b.
Propionibakterium-Infektion	≤2 oder n.b.	1 (4%) 2	≤2 oder n.b.
Reaktion im Zusammenhang mit einer Infusion	≤2 oder n.b.	1 (4%) 2	≤2 oder n.b.
Mindestens ein CTCAE Grad 3 SUE	8 (33%)	11 (46%)	4 (17%)
Mindestens ein CTCEA Grad 4 SUE	0	1 (4%)	0
Mindestens ein mit der Behandlung in Zusammenhang stehendes SUE	7 (29%)	8 (33%)	0
Mindestens ein SUE, welches zum Studienabbruch führte	0	0	0
Todesfälle	0	0	0
Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1. CSR 190-201: Tabelle 12.2.1.1, Tabelle 12.3.1.2.1, Tabelle 14.3.1.6.2 FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.1, Tabelle 14.3.1.12, Tabelle 14.3.1.6.1, Tabelle 14.3.1.6.2, Tabelle 14.3.1.17.1, Tabelle 14.3.1.17.2			

Bei 16 Patienten traten im Verlauf der Studie 190-201 insgesamt 32 SUEs auf. Bis zum Datenschnitt im Juni 2016 entwickelten 19 der 24 Patienten (79%) insgesamt 51 SUEs, davon wurde bei 8 Patienten (33%) ein Zusammenhang mit Cerliponase alfa angenommen. Lediglich die SUEs „Überempfindlichkeit“ (n=7, 29%), „Infektionen der oberen Atemwege“ (n=4, 17%), „Gastroenteritis“ (n=3, 13%), „Fieber“ (n=2, 8%), „Pharyngitis bakteriell“ (n=2, 8%) und „Epilepsie“ (n=2, 8%) traten bei mehr als einem Patienten auf. Bei 11 Patienten (46%) wurden drei SUEs mit CTCAE-Grad 3 berichtet, darunter Überempfindlichkeit Grad 3 bei zwei Patienten (8%). Bei jeweils einem Patienten (4%) traten folgende Grad 3-Ereignisse auf: Dysphagie, Clostridium difficile-Kolitis, Infektion im Zusammenhang mit einem medizinischen Gerät, Pharyngitis, Pharyngitis bakteriell, Pneumonie, Propionibacterium-Infektion, Hautinfektion, Azidose, intrakranielle Blutung, motorische Funktionsstörung, Krampfanfall, vaginaler Ausfluss, Schlafapnoe-Syndrom und Hypertrophie der Tonsillen. Zwei Patienten entwickelten eine Grad 3-Infektion der oberen Atemwege (8%; FDA Response: Tabelle 14.3.1.6.2). Ein schwerwiegender Fall von Status epilepticus CTCAE-Grad 4 wurde bei einem Patienten (4%) berichtet. Todesfälle und Studien-/Therapieabbrüche, die mit der Behandlung in Zusammenhang standen oder durch andere sicherheitsrelevante Gründe bedingt waren, traten nicht auf. Keiner der Patienten brach die Studie aufgrund eines SUEs ab (Tabelle 4-62).

### ***Inzidenz von CTCAE-Grad $\geq 3$ UEs***

Tabelle 4-63: Inzidenz der CTCAE Grad  $\geq 3$  UEs pro SOK („Sicherheits“-Population) [26,34]

<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Patienten mit mindestens einem Grad 3 UE	9 (38%)	12 (50%)	4 (17%)
Patienten mit mindestens einem Grad 4 UE	0	1 (4%)	0
Erkrankungen des Blutes und des Lymphsystems	1 (4%)	1 (4%)	0
Erkrankungen des Gastrointestinaltrakts	0	1 (4%)	0
Erkrankungen des Immunsystems	3 (13%)	3 (13%)	0
Überempfindlichkeit	3 (13%)	3 (13%)	0
Infektionen und parasitäre Erkrankungen	3 (13%)	6 (25%)	0
Infektion der oberen Atemwege	0	2 (8%)	1 (4%)
Stoffwechsel und Ernährungsstörungen	1 (4%)	2 (8%)	1 (4%)
Erkrankungen des Nervensystems	2 (8%)	3 (13%) (Grad 3) 1 (4%) (Grad 4)	0

<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Erkrankungen der Geschlechtsorgane und der Brustdrüse	1 (4%)	1 (4%)	0
Erkrankungen der Atemwege, des Brustraums und Mediastinums	2 (8%)	3 (13%)	0
Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1. CSR 190-201: Tabelle 14.3.1.2.3 FDA Response 190-202 Datenschnitt Juni 2016: Tabelle 14.3.1.2.3, Tabelle 14.3.1.13.3			

Bis zum Datenschnitt im Juni 2016 hatten 12 Patienten (50%) ein Grad 3-UE und ein Patient ein Grad 4-UE entwickelt. Mit Ausnahme von Überempfindlichkeit (n=3, 13%), trat keines der Grad 3-UEs bei mehr als einem Patienten auf. Diese Überempfindlichkeitsreaktionen waren die einzigen Grad 3-UEs, die nach Einschätzung der Prüfarzte mit Cerliponase in Zusammenhang standen. Ein Patient (4%) entwickelte einen Status epilepticus (Grad 4), dieser war aus Sicht des Prüfarztes allerdings nicht behandlungsbedingt (FDA Response: Tabelle 14.3.1.3.2).

### ***Studienabbruch aufgrund eines UEs***

Tabelle 4-64: Studienabbrüche aufgrund eines UE („Sicherheits“-Population) [26,34]

<b>Kategorie</b>	<b>Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Studienabbruch aufgrund eines UEs	0	0	0
Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. CSR 190-201: Tabelle 12.2.1.1 FDA Response 190-202 Datenschnitt Juni 2016: Tabelle 14.3.1.1			

Keiner der Patienten brach die Studie aufgrund eines UEs ab (Tabelle 4-64).

***UEs von besonderem Interesse: Inzidenz von Überempfindlichkeits- und anaphylaktischen Reaktionen***

Wie bei jedem Wirkstoff biologischen Ursprungs wurden auch bei Cerliponase alfa Überempfindlichkeitsreaktionen erwartet. Alle berichteten Überempfindlichkeitsreaktionen wurden durch die breite MedDRA SMQ „Überempfindlichkeit“ identifiziert und kein UE durch die SMQ „Anaphylaktische Reaktionen“. Beide Suchalgorithmen waren *a priori* festgelegt.

Tabelle 4-65: Präspezifizierte UEs besonderen Interesses: Inzidenz von Überempfindlichkeitsreaktionen („Sicherheits“-Population) [26,34]

Kategorie (SOK / PT)	Studie 190-201 n=24 Erhebung bis zum finalen Datenschnitt	Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016	Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016
Patienten mit mindestens einer Überempfindlichkeitsreaktion <sup>1</sup>	14 (58%) 24	14 (58%) 36	4 (17%) 7
Grad 1	7 (29%)	5 (21%)	1 (4%)
Grad 2	4 (17%)	6 (25%)	3 (13%)
Grad 3	3 (13%)	3 (13%)	0 (0%)
Grad 4	0 (0%)	0 (0%)	0 (0%)
Erkrankungen des Gastrointestinaltrakts	0	1 (4%) 1	1 (4%) 1
Stomatitis	0	1 (4%) 1	1 (4%) 1
Erkrankungen des Immunsystems	8 (33%) 14 Grad 3: 3 (13%)	10 (42%) 17 Grad 3: 3 (13%)	0
Überempfindlichkeit	8 (33%) 14 Grad 3: 3 (13%)	9 (38%) 15 Grad 3: 3 (13%)	0
Jahreszeitbedingte Allergie	0	1 (4%) 2	0
Infektionen und parasitäre Erkrankungen	2 (8%) 2	4 (17%) 5	2 (9%) 2
Konjunktivitis	2 (8%) 2	4 (17%) 5	2 (9%) 2
Erkrankungen der Haut und des Unterhautzellgewebes	5 (21%) 8	5 (21%) 13	2 (9%) 4
Dermatitis	0	1 (4%) 1	1 (4%) 1
Kontaktdermatitis	2 (8%) 3	2 (8%) 3	0
Ausschlag	2 (8%) 3	2 (8%) 7	2 (9%) 3
Urtikaria	2 (8%) 2	2 (8%) 2	0
<p><sup>1</sup> Überempfindlichkeitsreaktionen umfassten (1) die breite MedDRA SMQ „Überempfindlichkeit“ oder (2) den breiten SMQ-Algorithmus „Anaphylaktische Reaktionen“. Mittels des SMQ-Algorithmus „Anaphylaktische Reaktionen“ wurden keine UEs identifiziert.</p> <p>Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1.</p> <p>CSR 190-201: Tabelle 14.3.1.7.1, Tabelle 14.3.1.7.2</p> <p>FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.7.1, Tabelle 14.3.1.7.2, Tabelle 14.3.1.18.1</p>			

Bei 14 der 24 Patienten (58%) traten insgesamt 36 Überempfindlichkeitsreaktionen auf, diese waren allerdings bei nur drei Patienten schwer (CTCAE-Grad 3; n=3/24, 13%; Datenschnitt: Juni 2016). Bei den Grad 3-UEs bei drei Patienten handelte es sich per MedDRA PT um



Überempfindlichkeit. Konjunktivitis wurde bei 4/24 Patienten (17%) berichtet und bei jeweils zwei Patienten (8%) Kontaktdermatitis, Ausschlag und Urtikaria. Zwischen dem Abschlussbericht der Studie 190-201 und dem Datenschnitt im Juni 2016 wurden insgesamt 12 weitere Überempfindlichkeitsreaktionen berichtet (MedDRA SMQ), allerdings blieb die Zahl der Patienten mit Überempfindlichkeitsreaktionen unverändert (n=14/24, 58%) (Tabelle 4-65). Nach Einschätzung der Prüferärzte waren 9/15 Überempfindlichkeitsreaktionen (PT) bei 7/9 Patienten schwerwiegend (FDA Response: Tabelle 14.3.6.1). Insgesamt 14 dieser 15 Überempfindlichkeitsreaktionen (PT) traten bereits während der Studie 190-201 auf, darunter alle Grad 3-Ereignisse (Tabelle 4-65).

Alle Überempfindlichkeitsreaktionen, die während der Studie 190-201 aufgetreten waren, waren bis zum Studienende abgeklungen und es war in keinem Fall erforderlich, die Dosis des Prüfpräparates aufgrund des UEs zu verändern (CSR 190-201: Tabelle 12.3.1.3.2.1.3). Mit Ausnahme einer Überempfindlichkeitsreaktion, traf dies auch auf den Zwischenbericht der Studie 190-202 zu. Ein Fall von Heufieber (nicht behandlungsbedingt) war bis zum Zeitpunkt der Zwischenanalyse noch nicht vollständig angeklungen (CSR 190-202 Cut-off: 15.10.2015; Tabelle 12.3.1.3.2.2).

Wie in Abbildung 18 dargestellt, traten bei 10 Patienten (42%) keine, bei 7 Patienten (29%) eine, bei vier Patienten (17%) zwei und bei drei Patienten (13%) drei Überempfindlichkeitsreaktionen auf (Abschlussbericht der Studie 190-201).

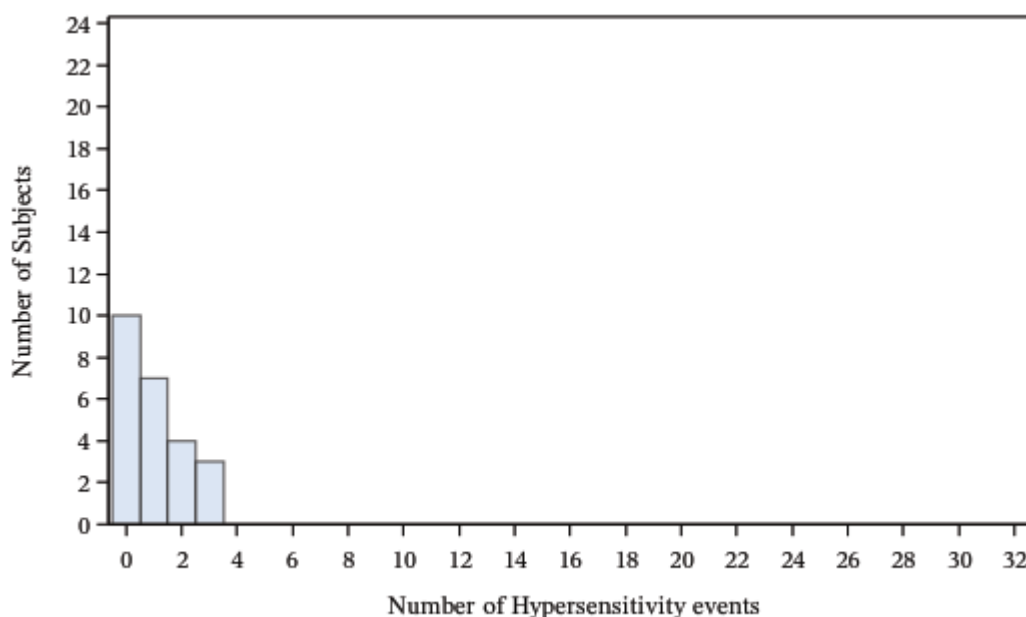


Abbildung 18: Anzahl der Überempfindlichkeitsreaktionen pro Patient („Sicherheits“-Population) [26]

CSR 190-201 (finaler Datenschnitt): Abbildung 14.3.1.7.9

Die Zahl der Patienten mit mindestens einer Überempfindlichkeitsreaktion (PT) schwankte nicht in Abhängigkeit von der Behandlungsdauer, sondern nahm im Studienverlauf tendenziell eher ab: >0-24 Wochen (n=4/24, 17%), >24-48 Wochen (n=5/23, 22%), >48-72 Wochen (n=1/23, 4%), >72-96 Wochen (n=1/23, 4%) und >96 Wochen (n=0/11) (FDA Response, Datenschnitt Juni 2016: Tabelle 14.3.1.2.9).

***UEs von besonderem Interesse: Inzidenz von UEs in zeitlichem Zusammenhang mit den Infusionen („temporally-related events“, TRE***

Tabelle 4-66: Präspezifizierte UEs besonderen Interesses: Inzidenz von UEs, die bei mindestens drei Patienten (auf SOK- und PT-Ebene) in einem zeitlichen Zusammenhang mit der Infusion von Cerliponase alfa auftraten („Sicherheits“-Population) [26,34]

<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Patienten mit mindestens einem UE, das in zeitlichem Zusammenhang mit der Infusion auftrat <sup>1</sup>	24 (100%)	24 (100%)	18 (78%)
Grad 1	4 (17%)	1 (4%)	7 (30%)
Grad 2	15 (63%)	16 (67%)	9 (39%)
Grad 3	5 (21%)	7 (29%)	2 (9%)
Grad 4	0 (0%)	0 (0%)	0 (0%)
IAR (SOK)			
Erkrankungen des Nervensystems	16 (67%)	19 (79%)	10 (43%)
Allgemeine Erkrankungen und Erkrankungen am Verabreichungsort	11 (46%)	11 (46%)	8 (35%)
Erkrankungen des Gastrointestinaltrakts	10 (42%)	12 (50%)	5 (22%)
Infektionen und parasitäre Erkrankungen	8 (33%) Grad 3: 1 (4%)	15 (63%) Grad 3: 2 (8%)	6 (26%) Grad 3: 1 (4%)
Erkrankungen des Immunsystems	8 (33%) Grad 3: 3 (13%)	9 (38%) Grad 3: 3 (13%)	0
Psychiatrische Erkrankungen	7 (29%)	7 (29%)	0
Erkrankungen des Blutes und Lymphsystems	3 (13%) Grad 3: 1 (4%)	3 (13%) Grad 3: 1 (4%)	0
Herzerkrankungen	2 (8%)	3 (13%)	1 (4%)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	4 (17%)	5 (21%)	2 (9%)
Untersuchungen	3 (13%)	4 (17%)	3 (13%)
Stoffwechsel und Ernährungsstörungen	3 (13%)	6 (25%) Grad 3: 1 (4%)	3 (13%) Grad 3: 1 (4%)

Kategorie (SOK / PT)	Studie 190-201 n=24 Erhebung zum finalen Datenschnitt	Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016	Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016
Produktprobleme	SOC nicht existent	5 (21%)	3 (13%)
IARs (PT)			
Fieber	11 (46%)	11 (46%)	6 (26%)
Überempfindlichkeit	8 (33%) Grad 3: 3 (13%)	9 (38%) Grad 3: 3 (13%)	0
Krampfanfall	8 (33%)	8 (33%)	3 (13%)
Erbrechen	5 (21%)	7 (29%)	2 (9%)
Tremor	3 (13%)	6 (25%)	0 (1 Grad 1 UE)
Problem mit einer Nadel	0	5 (21%)	0 (1 Grad 1 UE)
Obstipation	3 (13%)	4 (17%)	0 (1 Grad 1 UE)
Gangstörung	4 (17%)	4 (17%)	0 (1 Grad 1 UE)
Infektion der oberen Atemwege	< n=3	4 (17%)	0 (1 Grad 2 UE)
Kopfschmerz	3 (13%)	4 (17%)	0
Myoklonus	4 (17%)	4 (17%)	0
Diarrhö	< n=3	3 (13%)	2 (9%)
Gastroenteritis	< n=3	3 (13%)	2 (9%)
Sturz	< n=3	3 (13%)	0 (1 Grad 1 UE)
Generalisierte tonisch-klonische Anfälle	0	3 (13%)	2 (9%)
Infektion der Atemwege	< n=3	2 (8%)	0 (1 Grad 1 UE)
<p><sup>1</sup> Als mit der Infusion zeitlich in Zusammenhang stehende UEs waren UEs definiert, die nach Beginn der Infusion des Prüfpräparates sowie innerhalb von 24 Stunden nach Beginn oder Neubeginn der Infusion auftraten.</p> <p>Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 bzw. 19.1.</p> <p>CSR 190-201: Tabelle 14.3.1.8</p> <p>FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.8</p>			

Als TRE war jedes UE definiert, das nach Beginn der Infusion des Prüfpräparates bis zu 24 Stunden nach Beginn oder Neustart der Infusion aufgetreten war, unabhängig davon, ob ein Zusammenhang mit Cerliponase alfa bestand. Alle Patienten entwickelten mindestens ein TRE (100%; Datenschnitt Juni 2016). Die meisten TREs wurden innerhalb der SOKs „Erkrankungen des Nervensystems“ (79%), „Infektionen und parasitäre Erkrankungen“ (63%), „Erkrankungen des Gastrointestinaltrakts“ (50%), „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ (46%) und „Erkrankungen des Immunsystems“ (38%) berichtet. Die

häufigsten TREs nach PT waren Fieber (46%), Überempfindlichkeit (38%), Krampfanfall (33%) und Erbrechen (29%). Die meisten TREs waren leicht bis mittelschwer (CTCAE-Grade 1 oder 2). Sieben Patienten (29%) entwickelten über das gesamte Dosisintervall mindestens ein Grad 3-TRE (Tabelle 4-66).

***UEs von besonderem Interesse: Inzidenz von UEs, die das Nervensystem betreffen***

Status epilepticus, Hydrozephalus und Meningitis waren *a priori* als UEs von besonderem Interesse definiert. Zusätzliche UEs durch eine breite MedDRA-Abfrage „Konvulsionen“ (SMQ) identifiziert.

Tabelle 4-67: UEs besonderen Interesses: Inzidenz von UEs innerhalb der SOK „Erkrankungen des Nervensystems“ und Konvulsionen („Sicherheits“-Population, gesamtes Dosisintervall) [26,34]

<b>Kategorie (SOK / PT)</b>	<b>Studie 190-201 n=24 Erhebung zum finalen Datenschnitt</b>	<b>Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016</b>	<b>Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016</b>
Patienten mit mindestens einem UE innerhalb der SOK „Erkrankungen des Nervensystems“	24 (100%)	24 (100%)	17 (74%)
Grad 1	9 (38%)	8 (33%)	6 (26%)
Grad 2	13 (54%)	12 (50%)	11 (48%)
Grad 3	2 (8%)	3 (13%)	0 (0%)
Grad 4	0 (0%)	1 (4%)	0 (0%)
PT Hydrozephalus (AEOSI) <sup>2</sup>	0	0	0
PT Meningitis (AEOSI) <sup>2</sup>	0	0	0
Patienten mit mindestens einem therapiebedingten <sup>4</sup> UE innerhalb der SOK „Erkrankungen des Nervensystems“	17 (71%)	18 (75%)	1 (4%)
Patienten, bei denen mindestens ein UE innerhalb der SMQ Konvulsionen aufgetreten war <sup>1</sup>	23 (96%) 218 UEs	23 (96%) 362 UEs	13 (57%) 101
Grad 1	9 (38%)	7 (29%)	
Grad 2	14 (58%)	14 (58%)	
Grad 3	0 (0%)	1 (4%)	
Grad 4	0 (0%)	1 (4%)	
Patienten mit mindestens einem therapiebedingten <sup>4</sup> , durch die SMQ „Konvulsionen“ identifizierten UE	11 (46%) 21 UEs	12 (50%) 22 UEs	0
Mindestens ein TRE (SMQ „Konvulsionen“)	12 (50%) 42 UEs	15 (63%) 59 UEs	8 (35%) 11 UEs

Kategorie (SOK / PT)	Studie 190-201 n=24 Erhebung zum finalen Datenschnitt	Studie 190-201/202 n=24 Erhebung bis zum Datenschnitt: Juni 2016	Studie 190-202 n=23 Datenschnitt Juni 2016 Erhebung: 15.10.2015 bis 03.06.2016
Erkrankungen des Nervensystems (SMQ „Konvulsionen“)	23 (96%) 218 UEs	23 (96%) 362 UEs	13 (57%) 101 UEs
Atonische Anfälle	2 (8%) 4 UEs	2 (8%) 4 UEs	0
Klonischer Krampf	0	1 (4%) 1 UE	0
Drop attack <sup>3</sup>	2 (8%) 2 UEs	2 (8%) 2 UEs	0
Epilepsie Therapiebedingt <sup>4</sup>	11 (46%) 89 UEs 4 (17%) 4 UEs	11 (46%) 175 UEs 4 (17%) 4 UEs	5 (22%) 58 UEs 8 (35%) 11 UEs
Generalisierte tonisch-klonische Anfälle	2 (8%) 3 UEs	11 (46%) 27 UEs	9 (39%) 20 UEs
Epileptischer Myoklonus	1 (4%) 1 UE	1 (4%) 1 UE	0
Partielle Anfälle	2 (8%) 3 UEs	2 (8%) 3 UEs	0
Petit mal Epilepsie (AEOSI) <sup>2</sup>	2 (8%) 2 UEs	7 (29%) 13 UEs	3 (13%) 7
Krampfanfall Therapiebedingt <sup>4</sup>	14 (58%) 110 UEs 8 (33%) 13 UEs	14 (58%) 130 UEs 9 (38%) 14 UEs	7 (30%) 15 UEs 3 (13%) 3 UEs
Anfall-Cluster	2 (8%) 4 UEs	2 (8%) 5 UEs	1 (4%) 1
Status epilepticus (AEOSI) <sup>2</sup>	0	1 (4%) 1 UE	0
Patienten mit mindestens einem SUE Konvulsionen (SMQ)	2 (8%) 2 UEs	4 (17%) 4 UEs	0

<sup>1</sup> Entsprechende UEs wurden durch eine breite MedDRA-Abfrage identifiziert (SMQ „Konvulsionen“; Post-hoc-Analyse).

<sup>2</sup> Diese UEs (MedDRA PT) waren prospektiv als UEs besonderen Interesses definiert.

<sup>3</sup> Keine deutsche Übersetzung in EUTCT angegeben.

<sup>4</sup> UEs, die der Prüfarzt als mit der Behandlung in Zusammenhang stehend bewertete.

Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 (190-201) bzw. 19.1 (190-202).  
 CSR 190-201: Tabelle 14.3.1.2.3, Tabelle 14.3.1.3.1, Tabelle 14.3.1.9.1, Tabelle 14.3.1.9.2, Tabelle 14.3.1.9.3, Tabelle: 14.3.1.9.4, Tabelle 14.3.1.9.5  
 FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.3.2, Tabelle 14.3.1.9.1, 14.3.1.9.2, Tabelle 14.3.1.9.3, Tabelle 14.3.1.9.4, Tabelle 14.3.1.9.5, Tabelle 14.3.1.13.3, Tabelle 14.3.1.14.1, Tabelle 14.3.1.20.1, Tabelle 14.3.1.20.2, Tabelle 14.3.1.20.4, Tabelle 14.3.1.20.5

Alle Patienten entwickelten mindestens ein UE innerhalb der SOK „Erkrankungen des Nervensystems“, und diese waren mehrheitlich leicht bis mittelschwer (CTCAE-Grad 1 oder 2: n=20/24, 83%). Bei einem der 24 Patienten (4%, 1 UE) wurde Status epilepticus als UE berichtet und bei 7 Patienten (29%) insgesamt 13 Fälle von Petit mal Epilepsie. Es traten während der Studie 190-201/202 keine Fälle von Hydrozephalus und Meningitis auf, die den vorab definierten Kriterien entsprachen (Tabelle 4-67).

Durch eine breite MedDRA-Abfrage „Konvulsionen“ (SMQ) wurden bis zum Datenschnitt im Juni 2016 insgesamt 362 UEs bei 23/24 Patienten (96%) identifiziert, und dabei handelte es sich ausnahmslos um Erkrankungen des Nervensystems (SOK). Ausgenommen ein CTCAE Grad 3-UE (Krampfanfall) und ein Grad 4-UE (Status epilepticus, nicht behandlungsbedingt), waren alle, durch die MedDRA-Abfrage (SMQ) erfassten UEs von leichter bis mäßiger Schwere (Grad 1: n=7/24, 29%; Grad 2: n=14/24, 58%) (Tabelle 4-67). Bei vier der 24 Patienten (17%) waren vier Konvulsionsereignisse schwerwiegend, darunter zwei Fälle von Epilepsie (n=2/24, 8%) und bei jeweils einem Patienten ein Krampfanfall bzw. Status epilepticus (jeweils n=1, 4%; Datenschnitt Juni 2016: Tabelle 14.3.1.9.3). Zweiundzwanzig der insgesamt 362 UEs (22/362, 6%) bei 12/24 Patienten (50%) waren nach Einschätzung der Prüferärzte behandlungsbedingt. Insgesamt 59 der 362 UEs (16,3%) traten in zeitlichem Zusammenhang mit der Infusion von Cerliponase auf (TRE). Die häufigsten, durch die breite MedDRA-Abfrage erfassten UEs waren Epilepsie (175 UEs bei 11/24 Patienten, 46%) und Krampfanfälle (130 UEs bei 14/24 Patienten, 58%). Davon standen nur vier der insgesamt 175 Epilepsieereignisse (2,3%) und 14 der 130 Krampfanfälle (10,8%) aus Sicht der Prüferärzte mit der Behandlung in Zusammenhang (Tabelle 4-67). Sechs Patienten (25%) entwickelten insgesamt 7 Konvulsionen, die nach der Implantation des ICV-Ports aber vor der ersten Dosis Cerliponase alfa auftraten (CSR 190-201: S. 198).

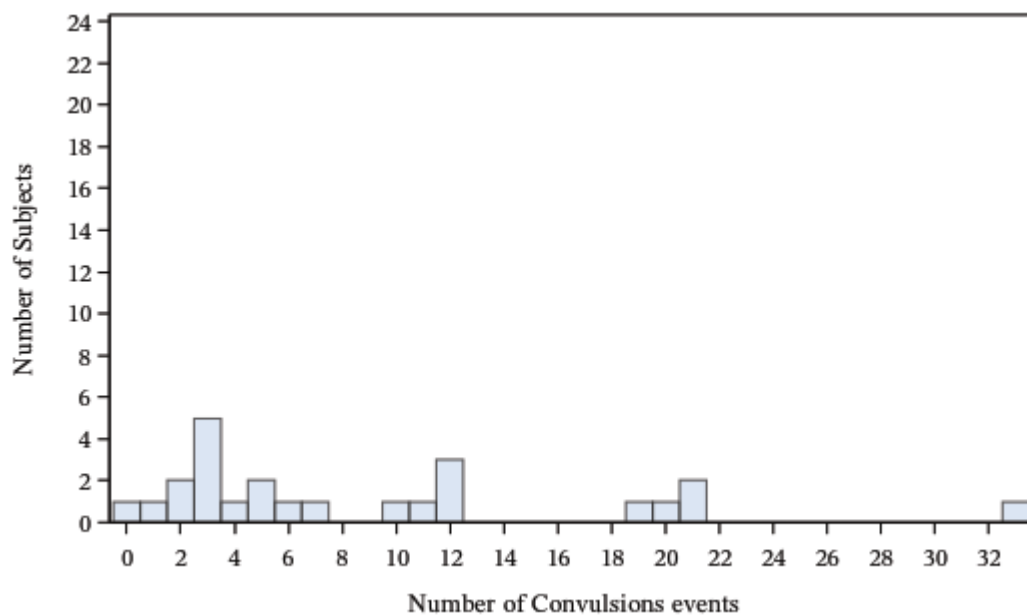


Abbildung 19: Anzahl der Patienten, bei denen eine oder mehrere Konvulsionen aufgetreten waren (SMQ „Konvulsionen“, „Sicherheits“-Population) [26]

CSR 190-201: Abbildung 14.3.1.9.6

Die Zahl der Patienten, die mindestens ein, das Nervensystem betreffendes UE entwickelten, nahm mit zunehmender Dauer der Behandlung tendenziell ab: >0-24 Wochen (n=23/24, 96%), >24-48 Wochen (n=19/23, 83%), >48-72 Wochen (n=18/23, 78%), >72-96 Wochen (n=12/23, 52%) und >96 Wochen (n=5/11, 45%). Mit der Dauer der Behandlung wurde Epilepsie zunehmend weniger häufig als UEs (MedDRA PT) berichtet: >0-24 Wochen (n=8/24, 33%),

>24-48 Wochen (n=8/23, 35%), >48-72 Wochen (n=6/23, 26%), >72-96 Wochen (n=4/23, 17%) und >96 Wochen (n=1/11, 9%). Auch hinsichtlich der Inzidenz von Krampfanfällen zeigte sich eine abnehmende Tendenz, je länger die Behandlung andauerte: >0-24 Wochen (n=11/24, 46%), >24-48 Wochen (n=10/23, 43%), >48-72 Wochen (n=8/23, 35%), >72-96 Wochen (n=2/23, 9%) und >96 Wochen (n=0/11).

***UEs von besonderem Interesse: Unerwartete und rasche Krankheitsprogression (gemessen als Abfall auf der CLN2-spezifischen ML-Skala), die nicht anders erklärbar ist***

Bei keinem der Patienten fiel der ML-Score unerwartet rasch und aus nicht anders erklärbaren Gründen ab (*a priori* definiertes UE von besonderem Interesse) (CSR 190-201: S. 198). Die Zwischenanalyse der Studie 190-202 bestätigt dies (15.10.2015).

***UEs in Zusammenhang mit dem ICV-Port***

UEs im Zusammenhang mit dem ICV-Port waren nicht definitiv *a priori* als UEs von besonderem Interesse definiert, durch den ICV-Zugang bedingte Komplikationen wurden allerdings schon bei der Studienplanung in Betracht gezogen.

Tabelle 4-68: UEs, die mit dem ICV-Zugang in Zusammenhang stehen („Sicherheits“-Population, gesamtes Dosisintervall) [26,34]

Kategorie (SOK / PT)	Studie 190-201 n=24 (Finaler Datenschnitt)	Studie 190-202 n=24 (Datenschnitt Juni 2016) n=24
Patienten mit mindestens einem mit dem ICV-Zugang in Zusammenhang stehendem UE <sup>1</sup>	7 (29%)	11 (46%) 30 UEs
Grad 1	6 (25%)	9 (38%)
Grad 2	0 (0%)	0 (0%)
Grad 3	1 (4%)	2 (8%)
Grad 4	0 (0%)	0 (0%)
Erkrankungen des Gastrointestinaltrakts	1 (4%)	1 (4%) 1 UE
Erbrechen	1 (4%)	1 (4%) 1 UE
Allgemeine Erkrankungen und Erkrankungen am Verabreichungsort	3 (13%)	2 (8%) 2 UEs
Problem eines Geräteanschlusses	1 (4%)	0
Apparatur-Leck	1 (4%)	0
Funktionsstörung eines Gerätes	1 (4%)	0
Komplikation bedingt durch ein medizinisches Gerät	1 (4%)	1 (4%) 1 UE <sup>2</sup>
Problem mit einer Nadel	1 (4%)	0
Schmerz	1 (4%)	1 (4%) 1 UE
Infektionen und parasitäre Erkrankungen	1 (4%)	2 (8%) 3 UEs
Infektion in Zusammenhang mit einem medizinischen Gerät	0	1 (4%) 1 UE (Grad 3)

Kategorie (SOK / PT)	Studie 190-201 n=24 (Finaler Datenschnitt)	Studie 190-202 n=24 (Datenschnitt Juni 2016) n=24
Propionibacterium-Infektion	1 (4%) (Grad 3)	1 (4%) 2 UEs (Grad 3)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	2 (8%)	2 (8%) 2 UEs
Hämatom nach Eingriff	1 (4%)	1 (4%) 1 UE
Wundkomplikation	1 (4%)	1 (4%) 1 UE
Untersuchungen	1 (4%)	2 (8%) 2 UEs
Liquoruntersuchung anomal	1 (4%)	1 (4%) 1 UE
Gefäßwiderstand systemisch erhöht	0	1 (4%) 1 UE
Stoffwechsel und Ernährungsstörungen	1 (4%)	1 (4%) 1 UE
Flüssigkeitsretention	1 (4%)	1 (4%) 1 UE
Erkrankungen des Nervensystems	3 (13%)	3 (13%) 3 UEs
Pleozytose	3 (13%)	3 (13%) 3 UEs
Produktprobleme <sup>3</sup>	0	7 (29%) 16 UEs
Problem eines Geräteanschlusses	0	1 (4%) 1 UE
Problem mit der Auslösung der Vorrichtung	0	1 (4%) 1 UE
Infusionsproblem der Vorrichtung	0	1 (4%) 1 UE
Apparatur-Leck	0	1 (4%) 5 UEs
Funktionsstörung eines Gerätes	0	2 (8%) 2 UEs
Verschluss eines medizinischen Gerätes	0	1 (4%) 1 UE
Problem mit einer Nadel	0	4 (17%) 5 UEs
<p><sup>1</sup> Die Prüfarzte bewerteten den Zusammenhang mit dem Medizinprodukt (ICV-Zugang).</p> <p><sup>2</sup> Kodiert als „Komplikation in Verbindung mit dem Gerät“.</p> <p><sup>3</sup> Die SOK wurde erst nach der MedDRA-Version 18.1 (190-201) eingeführt.</p> <p>Bei Patienten, bei denen mehr als 1 UE innerhalb einer MedDRA SOK oder ein PT mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt. Kodiert nach MedDRA 18.1 (190-201) und 19.1 (190-202).</p> <p>CSR 190-201: Tabelle 14.3.1.4.1, Tabelle 14.3.1.4.2</p> <p>FDA Response Datenschnitt Juni 2016: Tabelle 14.3.1.4.1, Tabelle 14.3.1.4.3</p>		

Bei 11 der 24 Patienten (46%) traten UEs auf, die nach Einschätzung der Prüfarzte mit dem ICV-Port in Zusammenhang standen (Datenschnitt: Juni 2016). Diese UEs waren bei 9 Patienten nur leichter Ausprägung (CTCAE-Grad 1: 38%) und bei zwei Patienten (8%) traten Grad 3-UEs (Infektionen) auf. Lediglich die UEs „Problem mit einer Nadel“ (5 UEs bei 4 Patienten, 17%), Pleozytose (3 UEs bei 3 Patienten, 13%; Surrogatparameter von unbekannter Patientenrelevanz) und Apparatur-Leck (5 UEs bei 1 Patient, 4%) wurden während der Studie 190-201/202 mehr als einmal berichtet (Tabelle 4-68). Bei zwei Patienten musste der ICV-Port während der Studie 190-201/202 ersetzt werden und zwar in beiden Fällen aufgrund einer



schwerwiegenden, Grad 3-Infektion: *Propionibacterium*-Infektion bzw. Infektion mit *Staphylococcus epidermidis* in Zusammenhang mit dem ICV-Port. Beide UEs klangen unter Behandlung mit Antibiotika und nach Entfernen des ICV-Ports ab. Beide Patienten setzten die Studie fort und vertrugen die nachfolgenden Cerliponase alfa-Infusionen (CSR 190-202: S. 168-169).

Hinsichtlich der Inzidenz von UEs im Zusammenhang mit dem ICV-Port zeigte sich im Studienverlauf bzw. abhängig von der Behandlungsdauer kein klarer Trend: >0-24 Wochen n=4/24 (17%), >24-48 Wochen: n=3/23 (13%), >48-72 Wochen: n=6 (26%), >72-96 Wochen: n=3/23 (13%), >96 Wochen: n=1/11 (9%) (Datenschnitt Juni 2016: Tabelle 14.3.1.4.8). Pro 1.000 Cerliponase alfa-Infusionen (gesamt 1.174 Infusionen bis Juni 2016) wurden 2,56 UEs im Zusammenhang mit dem ICV-Port berichtet (Datenschnitt Juni 2016: Tabelle 14.3.1.4.9).

### ***Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext***

Die Mehrzahl der Patienten wurde in dem deutschen Studienzentrum rekrutiert (n=12/24). Sollten sich die Begleitmedikationen oder andere Therapiemaßnahmen zwischen den Studienzentren unterscheiden, kann dies einen Einfluss auf die Sicherheitsanalyse haben. Patienten mit CLN2 erhalten in der Regel eine Vielzahl symptomatischer oder palliativer Therapien, allerdings gibt es bislang keinen Behandlungsstandard oder entsprechende Richtlinien. Insgesamt 24 CLN2-Experten, darunter auch Patientenvertreter, nahmen weltweit an einer Online-Umfrage teil, um Einblicke in das aktuelle Krankheitsmanagement zu gewinnen. Demnach erhalten die Patienten zahlreiche Medikamente zur symptomatischen Behandlung von Krampfanfällen, Myoklonus, Spastik, Dystonie, vermehrtem Sekretfluss und Schmerzen [40]. Da die Behandlung von Patienten mit CLN2 grundsätzlich nur erfahrenen Klinikern vorbehalten ist, erscheint es unwahrscheinlich, dass sich die symptomatische/palliative Therapie zwischen den Teilnehmern der Studie 190-201/202 und Patienten, die unter üblichen Praxisbedingungen in Deutschland behandelt werden, nennenswert unterscheidet.

Da nur wenige Daten bei unbehandelten Patienten vorliegen, kann nicht vollständig ausgeschlossen werden, dass sich z. B. die Behandlungsstrategien oder die Art, wie UEs berichtet werden, zwischen den Ländern geringfügig, allerdings wahrscheinlich nicht erheblich unterscheiden (insbesondere in Anbetracht der Tatsache, dass die Studie ausschließlich in westlichen Ländern mit erwartungsgemäß ähnlichen Gesundheitssystemen stattfindet). Die häufigsten UEs, die in dem deutschen Zentrum berichtet wurden (n=12 Patienten), betrafen die SOCs „Erkrankungen des Gastrointestinaltrakts“ (75%), „Erkrankungen des Immunsystems“ (83%; darunter „Überempfindlichkeit“ [PT] mit 75% am häufigsten), „Infektionen und parasitäre Erkrankungen“ (100%) und „Erkrankungen des Nervensystems“ (100%; darunter Epilepsie [PT] mit 92% am häufigsten) (FDA Response Juni 2016: Tabelle 14.3.1.2.7). Diese stellen typische UEs dar, die bei Kleinkindern/Kindern oder im Rahmen einer ERT zu erwarten sind, oder sind charakteristische Manifestationen der Erkrankung. Daher entspricht das Sicherheitsprofil von Cerliponase alfa in dem deutschen Zentrum demjenigen, welches bei schwerkranken Kleinkindern und Kindern auch angesichts der zahlreichen Begleitmedikationen zu erwarten ist.

#### 4.4.2.3.4 Subgruppenanalysen – weitere Untersuchungen

##### *Vergleich von Cerliponase alfa (Studie 190-201/202) vs. Studie 190-901 (historische Kontrolle mit Paarbildung, ISE) [57]*

Entsprechend ISE SAP waren keine Subgruppenanalysen vorgesehen.

##### *Studie 190-201/202 (ohne Paarbildung)*

Der SAP sah fakultativ bestimmte Subgruppenanalysen vor, um den Behandlungseffekt näher zu charakterisieren. Es wurde darauf verzichtet, den Einfluss verschiedener Subgruppenmerkmale wie Alter, Geschlecht oder ethnische Herkunft auf den Behandlungseffekt formal, mittels statistischer Verfahren zu untersuchen, da die Ergebnisse auf Basis kleiner Subgruppen bei einer Gesamtfallzahl von nur 23 auswertbaren Patienten kaum interpretierbar wären. Zu Beginn der Studie 190-201 (Baseline) waren 20 Patienten unter und vier Patienten über fünf Jahre alt. Insgesamt 15 der 24 Patienten (63%) waren weiblich und mit Ausnahme eines Asiaten, waren alle Patienten weiß (CSR 190-201/202).

Im Rahmen einer prospektiv geplanten Subgruppenanalyse wurde der Abfall des ML-Scores über 48 Wochen getrennt nach Studienzentrum untersucht. Der mittlere ML-Score (SD) schwankte zu Baseline zwischen 3,2 (0,75; Italien, n=6) und 4,3 Punkten (0,58; USA, n=3) und lag in Deutschland mit 3,5 Punkten (1,24; n=12) im unteren Bereich. Der ML-Score (Mittelwert  $\pm$  SD in Punkten) fiel über 48 Wochen in den einzelnen Studienzentren wie folgt ab (300 mg-Dosisintervall; Ausschluss von zwei Patienten mit ML-Score stabil = 6): Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany (Angela Schulz, n=10): 0,11 (0,660); Bambino Gesù Children's Hospital, Rom, Italien (Nicola Specchio; n=6): 0,83 (0,983); Nationwide Children's Hospital, Columbus, Ohio, USA (Emily de los Reyes, n=3): 0,67 (1,155); Great Ormond Street Hospital, London, UK (Paul Gissen, n=2): 0,50 (0,707). Angesichts der kleinen Fallzahlen pro Studienzentrum sollten diese Ergebnisse nur mit Vorsicht interpretiert werden (Tabelle 4-69).

Tabelle 4-69: Geschätzte Rate des Abfalls des ML-Scores (Punkte pro 48 Wochen) nach Studienzentrum („Wirksamkeits“-Population, 300 mg-Dosisintervall) [26]

Rate des Abfalls (Punkte/48 Wochen)	Schulz (n=10) Germany	Specchio (n=6) Italy	de los Reyes (n=3) USA	Gissen (n=2) UK	Williams (n=1) UK
<i>CSR 190-201</i>					
Mittelwert (SD)	0,11 (0,660)	0,83 (0,983)	0,67 (1,155)	0,50 (0,707)	Aus „Wirksamkeits“-Population ausgeschlossen
95% KI	-0,36; 0,58	-0,20; 1,86	-2,20; 3,54	-5,85; 6,85	
Positive Werte entsprechen einer klinischen Progression und negative Werte einer Verbesserung des klinischen Zustands. Berechnung des Abfalls über 48 Wochen: (ML Score zu Baseline – letzter ML-Score)/(48 Wochen) CSR 190-201: Tabelle 14.2.2.4					

#### 4.4.2.4 Zusammenfassung der Ergebnisse aus weiteren Unterlagen

Cerliponase alfa (Brineura®) ist zur Behandlung einer seltenen Erkrankung indiziert und gemäß Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 als Arzneimittel zur Behandlung seltener Leiden („Orphan Drug“) zugelassen [14]. Für Arzneimittel zur Behandlung eines seltenen Leidens gilt gemäß § 35a SGB V Abs. 1 Satz 10 der medizinische Zusatznutzen durch die Zulassung als belegt [15].

#### *Efficacy (morbidity) Wirksamkeit (Morbidity)*

Tabelle 4-70 gibt einen Überblick der Ergebnisse der patientenrelevanten Endpunkte, die zur Bewertung der Wirksamkeit von Cerliponase alfa herangezogen werden (Datenschnitt: 15. Oktober 2015).

Tabelle 4-70: Zusammenfassung der Ergebnisse patientenrelevanter Wirksamkeits-Endpunkte

Endpunkt	190-201/190-901 (1:1 Paarbildung)			Bewertung des Effekts
	Effektmaß	Effekt	p-Wert	
<i>ML/HML Skala / Primäre Analysen (siehe Kapitel 4.4.2.3.1.3)</i>				
Zeit bis zur ersten stabilen Abnahme des ML/HML-Scores um $\geq 2$ Punkte oder ML = 0	HR, Kaplan-Meier	0,10 95% KI 0,03; 0,38	0,0005	Schwerwiegendes Symptom, 10-fach höheres Risiko einer Krankheitsprogression bei unbehandelten Patienten; erheblicher Zusatznutzen
Responder (Abnahme des ML/HML-Scores um $< 2$ Punkte/48 Wochen)	RR	2,00 95% KI 1,2420; 3,2207	0,00435	Schwerwiegendes Symptom, RR (iRR) $> 5$ („dramatischer“ Effekt); erheblicher Zusatznutzen
Non-Responder (stabile Abnahme des ML/HML-Scores um $\geq 2$ Punkte/48 Wochen)	RR	0,16667 95% KI 0,04212; 0,65942 (iRR: 6,00)	0,01067	
<i>Einzeldomänen (Motorische Fähigkeiten, Sprachvermögen) (siehe Kapitel 4.4.2.3.1.5)</i>				
Responder (Abnahme in der motorischen Domäne um $< 1$ Punkt/48 Wochen)	RR	1,900 95% KI 1,1676; 3,0919	0,0098	
Non-Responder (stabile Abnahme in der motorischen Domäne um $\geq 1$ Punkt/48 Wochen)	RR	0,250 95% KI 0,0817; 0,7652 (iRR: 4,00)	0,0151	
Responder (Abnahme in der Sprachdomäne um $< 1$ Punkt/48 Wochen)	RR	3,800 95% KI 1,7461; 8,2696	0,0008	

Endpunkt	190-201/190-901 (1:1 Paarbildung)			Bewertung des Effekts
	Effektmaß	Effekt	p-Wert	
Non-Responder (stabile Abnahme in der Sprachdomäne um $\geq 1$ Punkt/48 Wochen)	RR	0,1250 95% KI 0,327; 0,4773 (iRR: 8,00)	0,0023	
iRR: inverses relatives Risiko; KI: Konfidenzintervall; RR: relatives Risiko Kriterien zur Paarbildung (matching): Identischer ML/HML-Skalenwert und Altersunterschied $\leq 12$ Monate zu Baseline (300 mg-Baseline in Studie 190-201).				

### Vergleich der Studienpopulation (190-201) mit der historischen Kontrolle („matched“)

Der primäre Endpunkt der ISE war eine Responder-Analyse (als Responder bzw. Non-Responder galten Patienten, deren Erkrankung nicht fortgeschritten bzw. fortgeschritten war, ausgedrückt als Abfall des ML/HML-Scores um  $< 2$  bzw.  $\geq 2$  Punkte/48 Wochen). Zusätzlich war im SAP eine Ereigniszeitanalyse *a priori* festgelegt (Zeit bis zum ersten, nachhaltigen Abfall des ML/HML Scores um  $\geq 2$  Punkte oder bis zum ersten Wert = 0, d. h. vollständiger Funktionsverlust, der bis zum Behandlungsende erhalten blieb).

In die primäre Analyse wurden die Patienten der ITT-Population der Studie 190-201/202 eingeschlossen, denen nach zuvor festgelegten Kriterien Patienten der historischen Kontrollgruppe (190-901) im Verhältnis 1:1 zugeteilt wurden („matching“). Voraussetzung für die Paarbildung war, dass der ML/HML-Score exakt und das Alter (jeweils zu Baseline) soweit wie möglich (aber Altersunterschied nicht mehr als 12 Monate) übereinstimmten.

In der primären Responder-Analyse betrug das RR=2,00 ([95% KI: 1,2420; 3,2207], p=0,0044) zugunsten der Behandlung mit Cerliponase alfa. Hinsichtlich des Anteils der Patienten, die nicht auf die Therapie angesprochen hatten (Non-Responder), war das RR=0,167 ([95% KI: 0,04212; 0,65942], p=0,01067). Diese Ergebnisse belegen hinsichtlich der Krankheitsprogression einen großen Unterschied zwischen den behandelten und unbehandelten Patienten. Verschiedene Sensitivitätsanalysen unter Einsatz unterschiedlicher Algorithmen zur Paarbildung und Analysepopulationen bestätigen die Primäranalyse (Tabelle 4-42, Tabelle 4-43, Tabelle 4-44).

Bezogen auf die motorische Domäne der ML/HML-Skala, betrug die Ansprechrate (Abfall um  $< 1$  Punkt innerhalb von 48 Monaten) 86% (n=19/22) in der Cerliponase alfa-Gruppe verglichen mit 45% (n=10/22) in der historischen Kontrollgruppe („matched“; Tabelle 4-45). Das entsprechende RR (Responder) lag bei 1,900 ([95% KI 1,1676. 3,0919], p=0,0098) zugunsten von Cerliponase alfa. Das RR für Non-Response betrug 0,250 ([95% KI 0,0817, 0,7652], p=0,0151) und das Ergebnis war ebenfalls hochsignifikant zugunsten von Cerliponase alfa. Bei 90% (n=19/21) der Patienten der Cerliponase alfa-Gruppe und 24% (n=5/21) der Patienten der historischen Kontrollgruppe verschlechterte sich die Sprachdomäne der ML-Skala über 48 Wochen um weniger als einen Punkt (diese galten analog zur motorischen Domäne als

Responder; Tabelle 4-45). Das entsprechende RR für Response betrug hinsichtlich der sprachlichen Fähigkeiten 3,800 ([95% KI 1,7461; 8,2696],  $p=0,0008$ ) und für Non-Response 0,1250 ([95% KI 0,0327; 0,4773],  $p=0,0023$ ) (Tabelle 4-45, Tabelle 4-70). Der größere Behandlungseffekt von Cerliponase alfa, verglichen mit der historischen Kontrolle, auf die sprachlichen Fähigkeiten erklärt sich dadurch, dass mehr Patienten der historischen Kontrolle sprachliche als motorische Probleme entwickelten. Allerdings verbesserte Cerliponase alfa sowohl die sprachlichen als auch die motorischen Fähigkeiten vergleichbar, und beide Domänen trugen zu dem Effekt auf die kombinierte ML/HML-Skala bei.

In einer *a priori* definierten Kaplan-Meier-Analyse, betrug die HR der Krankheitsprogression (gemessen als Abfall um zwei Punkte auf der ML/HML-Skala oder  $ML/HML = 0$ ) 0,10 ([95%KI 0,03; 0,38],  $p=0,0005$ ) zugunsten von Cerliponase alfa verglichen mit der historischen Kontrolle („matched“). Der Behandlungseffekt ist so groß, dass er als „dramatisch“ eingestuft werden kann. Verschiedene Sensitivitätsanalysen bestätigen die primäre Ergebniszeitanalyse (Tabelle 4-47). Follow-up-Analysen bei Patienten, die über mindestens 96 Wochen beobachtet wurden, zeigen, dass der Behandlungseffekt im Studienverlauf sogar noch weiter zunimmt.

Zusätzliche Analysen zum Abfall der CLN2-spezifischen Skalen werden nachfolgend kurz zusammengefasst (diese Endpunkte werden allerdings nicht zur Quantifizierung des Zusatznutzens herangezogen):

Der Abfall des HML/ML-Scores (in Punkten über 48 Wochen) wurde bei der ITT-Population der Studie 190-201/202 und den auswertbaren Patienten der Studie 190-901 nach 1:1 Paarbildung (Methode #1) berechnet. Der HML/ML-Score (in Punkten/48 Wochen) fiel bei den Patienten der Studie 190-901 im Mittel (SD) um 2,06 (1,379) Punkte ab verglichen mit 0,53 (0,737) Punkten bei den Patienten der Studie 190-201/202, was einem Unterschied von 1,53 Punkten/48 Wochen ( $p<0,0001$ ) entspricht. Die verschiedenen, unterstützenden und Sensitivitätsanalysen zeigten übereinstimmend einen ähnlich großen Unterschied zwischen den behandelten und unbehandelten Patienten.

Die Analyse der Einzeldomänen (motorisch und Sprache, Abfall der Scores im Studienverlauf) bestätigt den Behandlungseffekt von Cerliponase alfa, der sich bereits bei der Analyse der kombinierten ML/HML-Skala gezeigt hatte. Der mittlere Gruppenunterschied (SE) betrug hinsichtlich der motorischen Funktion 0,72 (0,240) Punkte/48 Wochen ( $p=0,0055$ ) und 1,01 (0,200) Punkte/48 Wochen ( $p<0,0001$ ; Tabelle 4-49) hinsichtlich der sprachlichen Fähigkeiten.

Zusätzlich wurde der Summenscore, bestehend aus den Domänen motorische Funktion, Sprache und Sehen, analysiert bzw. dessen Verlauf (MLV/HMLV-Score: 0-9 Punkte). Bei den auswertbaren Patienten der Studie 190-901 ( $n=39$ ) nahm der MLV/HMLV-Score im Mittel (SD) um 2,23 (1,152) Punkte/48 Wochen ab verglichen mit 0,78 (1,042) Punkten/48 Wochen bei den Patienten der Studie 190-201/202 ( $n=21$ , „Wirksamkeits“-Population). Der Gruppenunterschied betrug 1,44 Punkte/48 Wochen ( $p<0,0001$ ) (Tabelle 4-50).

Der CLN2-Gesamt-Score (ML plus Sehen und Krampfanfälle) betrug bei den 37 Patienten der Studie 190-901 in Woche 49 7,0 (SD 2,41) Punkte und bei den 23 Patienten der Studie 190-

201/202 7,9 (SD 2,08) Punkte. Nach 48 Wochen war der Summenscore (Mittelwert  $\pm$  SD), bezogen auf Baseline, bei den unbehandelten Patienten um 2,2 (1,85) Punkte abgefallen verglichen mit -0.1 (1.94) Punkten bei den behandelten Patienten (Tabelle 4-51).

### Lebensqualität

#### PedsQL™

Tabelle 4-71: Überblick der Ergebnisse zur Lebensqualität (PedsQL™; ITT, n=23)

Endpunkt	Änderung bis zur letzten Erhebung (Punkte), Mittelwert (SD)
PedsQL™ Fremdbericht bei Kleinkindern (generischer Gesamtscore; Tabelle 4-54)	2,6 (12,16)
Körperliche Fähigkeiten	-6,1 (21,72)
Emotionale Fähigkeiten	7,3 (26,22) (n=22)
Soziale Fähigkeiten	7,8 (16,22)
Schule	6,3 (29,82)
Psychosozialer Summenscore	7,9 (14,91)
PedsQL™ „Family Impact Module“, Fremdbericht (Tabelle 4-54)	3,7 (19,04)
Körperliche Funktion	1,6 (20,86)
Emotionale Fähigkeiten	10,2 (22,99)
Soziale Fähigkeiten	-4,6 (25,23)
Kognitive Fähigkeiten	7,4 (32,82)
Kommunikative Fähigkeiten	2,2 (28,00)
Summenscore (körperlich, emotional, sozial und kognitiv)	4,0 (19,10)
Ein Score von 0 entspricht dem schlechtesten Wert und ein Score von 100 dem besten Wert.	

Der Mittelwert (SD) des PedsQL™-Moduls „Parent Report for Toddlers“ nahm zwischen Baseline und Woche 49 der stabilen Dosisphase um 2,6 (12,16) Punkte zu, d. h. der Wert verbesserte sich um etwa 4,3% (ITT, gesamtes Dosisintervall).

Der Mittelwert (SD) des PedsQL™-Moduls „Family Impact“ stieg zwischen Baseline und Woche 49 der stabilen Dosisphase um 3,7 (19,04) Punkte an, d. h. verbesserte sich um circa 6,0%. Insgesamt änderten sich die einzelnen Domänen des PedsQL™ im Studienverlauf von der ersten (Baseline) bis zur letzten Beobachtung nur wenig. Nur die Werte zur körperlichen Leistungsfähigkeit (Domäne „physical function“; „Parent Report for Toddlers“) und gesellschaftlichen Funktion (Domäne „social function“; „Family Impact“) nahmen bis zur letzten Beobachtung geringfügig ab (Tabelle 4-71). Angesichts der besonderen Schwere der Erkrankung ist die numerische Verbesserung, die sich bei den Gesamt-Scores beider Module zeigt, bemerkenswert. Diese positiven Resultate unterstützen die dramatischen Wirksamkeitsergebnisse nachdrücklich, die unter Behandlung mit Cerliponase beobachtet werden.

Tabelle 4-72: Überblick der Ergebnisse zur Lebensqualität (CLN2-spezifisches QoL-Instrument; ITT, n=23)

Endpunkt	Änderung bis zur letzten Erhebung (Punkte), Mittelwert (SD)
CLN2-spezifische Lebensqualität, Gesamtscore (Tabelle 4-56)	8,1 (14,33)
Krampfanfälle	18,2 (23,02)
Ernährung (ohne Gastrostomiesonde)	9,8 (22,15) (n=21)
Schlaf	4,1 (19,74)
Verhalten	7,2 (18,01)
Alltagsaktivitäten	-0,8 (18,10)
Ein Score von 0 entspricht dem schlechtesten Wert und ein Score von 100 dem besten Wert.	

Übereinstimmend mit den Ergebnissen der PedsQL™-Instrumente, nahm auch der CLN2-spezifische QoL-Score zwischen Baseline und Woche 49 der stabilen Dosisphase im Mittel (SD) um 8,1 (14,33) Punkte zu, was einer Verbesserung um etwa 10,9% entspricht. Die Behandlung mit Cerliponase alfa hatte keinen negativen Effekt auf irgendeine der Einzeldomänen. Mit Ausnahme der Domäne Alltagsaktivitäten (vernachlässigbarer Abfall) verbesserten sich alle Domänen und insbesondere die Krampfneigung der Patienten (Tabelle 4-56).

### ***Verträglichkeit und Sicherheit***

Auf Grundlage der Empfehlungen des IQWiGs werden UEs, die den Nutzen einer Intervention teilweise oder vollständig aufheben könnten, und solche, die mit schwerwiegender Morbidität oder sogar Mortalität verbunden sind oder die Lebensqualität deutlich verschlechtern, als patientenrelevant erachtet [23]. Bei einer sehr schweren und letztlich fatalen Erkrankung, für die keine krankheitsmodifizierende Therapie zur Verfügung steht, können nur schwere (CTCAE-Grad  $\geq 3$ ), schwerwiegende oder lebensbedrohliche UEs, UEs, die zum Therapieabbruch oder dauerhaften Beeinträchtigungen führen, das Risiko-Nutzen-Verhältnis beeinflussen. Weniger schwere UEs werden bei der Bewertung des Zusatznutzens von Cerliponase alfa nicht berücksichtigt, da diese in Übereinstimmung mit dem Addendum des CONSORT-Statements [54] das Verhältnis zwischen Nutzen und Schaden nicht verändern können, insbesondere in Anbetracht der besonderen Schwere der Erkrankung ohne Therapieoptionen.

Tabelle 4-73: Überblick der in der Studie 190-201/202 aufgetretenen unerwünschten Ereignisse (behandlungsspezifische UEs, n=24, „Sicherheits“-Population)

UEs nach MedDRA SOK oder PT	Inzidenz, n (%)	
	Studie 190-201	Studie 190-201/202 (Datenschnitt Juni 2016)
Patienten mit mindestens einem UE (Tabelle 4-59)	24 (100%)	24 (100%)
Inzidenz von UEs nach SOK ( $\geq 50\%$ ; Tabelle 4-60)		
Erkrankungen des Nervensystems	24 (100%)	24 (100%)
Infektionen und parasitäre Erkrankungen	23 (96%)	24 (100%)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	16 (67%)	19 (79%)
Erkrankungen des Gastrointestinaltrakts	19 (79%)	19 (79%)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	17 (71%)	19 (79%)
Psychiatrische Erkrankungen	12 (50%)	12 (50%)
Sehr häufige UEs ( $\geq 30\%$ , Tabelle 4-61)		
Fieber <sup>2</sup>	13 (54%)	17 (71%)
Krampfanfall Grad 3	14 (58%)	14 (58%) 1 (4%)
Erbrechen	11 (46%)	12 (50%)
Infektion der oberen Atemwege	10 (42%)	12 (50%)
Epilepsie	11 (46%)	11 (46%)
Generalisierte tonisch-klonische Anfälle	n.b. oder $<10\%$	11 (46%)
Rhinitis	5 (21%)	10 (42%)
Nasopharyngitis	7 (29%)	9 (38%)
Überempfindlichkeit Grad 3	8 (33%) 3 (13%)	9 (38%) 3 (13%)
Obstipation	7 (29%)	8 (33%)
Virusinfektion	5 (21%)	8 (33%)
Inzidenz von SUEs oder fatalen UEs (Tabelle 4-62)		
Mindestens ein SUE	16 (67%) 32	19 (79%) 51
SUEs, die bei mindestens zwei Patienten oder mehrmals auftraten		
Überempfindlichkeit	6 (25%)	7 (29%) 9
Infektion der oberen Atemwege	$\leq 2$ oder n.b.	4 (17%) 4
Gastroenteritis	$\leq 2$ oder n.b.	3 (13%) 3
Fieber	2 (8%)	2 (8%) 2



UEs nach MedDRA SOK oder PT	Inzidenz, n (%)	
	Studie 190-201	Studie 190-201/202 (Datenschnitt Juni 2016)
Pharyngitis bakteriell	2 (8%)	2 (8%) 3
Epilepsie	2 (8%)	2 (8%) 2
Propionibakterium-Infektion	≤2 oder n.b.	1 (4%) 2
Reaktion im Zusammenhang mit einer Infusion	≤2 oder n.b.	1 (4%) 2
Mindestens ein CTCAE Grad 3 SUE	8 (33%)	11 (46%)
Mindestens ein CTCAE Grad 4 SUE	0	1 (4%)
Todesfälle	0	0
Inzidenz von CTCAE Grad ≥3 UEs (>1 Patient pro SOK oder PT; Tabelle 4-63)		
Patienten mit mindestens einem Grad 3 UE	9 (38%)	12 (50%)
Patienten mit mindestens einem Grad 4 UE	0	1 (4%)
Erkrankungen des Immunsystems	3 (13%)	3 (13%)
Überempfindlichkeit	3 (13%)	3 (13%)
Infektionen und parasitäre Erkrankungen	3 (13%)	6 (25%)
Infektion der oberen Atemwege	0	2 (8%)
Stoffwechsel und Ernährungsstörungen	1 (4%)	2 (8%)
Erkrankungen des Nervensystems	2 (8%)	3 (13%) (Grad 3) 1 (4%) (Grad 4)
Erkrankungen der Atemwege, des Brustraums und Mediastinum	2 (8%)	3 (13%)
Studienabbruch aufgrund eines UEs (Tabelle 4-64)	0	0
UEs besonderen Interesses		
Überempfindlichkeitsreaktionen inkl. der häufigsten SOK (Tabelle 4-65)		
Patienten mit mindestens einer Überempfindlichkeitsreaktion (MedDRA SMQ)	14 (58%) 24 UEs	14 (58%) 36 UEs
Erkrankungen des Immunsystems	8 (33%) 14 UEs Grad 3: 3 (13%)	10 (42%) 17 UEs Grad 3: 3 (13%)
Überempfindlichkeit (PT)	8 (33%) 14 UEs Grad 3: 3 (13%)	9 (38%) 15 UEs Grad 3: 3 (13%)
Patienten mit mindestens einem UE, das in zeitlichem Zusammenhang mit der Infusion auftrat (IAR; Tabelle 4-66)	24 (100%)	24 (100%)
IAR bei ≥30% der Patienten (SOK bzw. PT)		
Erkrankungen des Nervensystems	16 (67%)	19 (79%)

UEs nach MedDRA SOK oder PT	Inzidenz, n (%)	
	Studie 190-201	Studie 190-201/202 (Datenschnitt Juni 2016)
Allgemeine Erkrankungen und Erkrankungen am Verabreichungsort	11 (46%)	11 (46%)
Erkrankungen des Gastrointestinaltrakts	10 (42%)	12 (50%)
Infektionen und parasitäre Erkrankungen	8 (33%) Grad 3: 1 (4%)	15 (63%) Grad 3: 2 (8%)
Erkrankungen des Immunsystems	8 (33%) Grad 3: 3 (13%)	9 (38%) Grad 3: 3 (13%)
Fieber	11 (46%)	11 (46%)
Überempfindlichkeit	8 (33%) Grad 3: 3 (13%)	9 (38%) Grad 3: 3 (13%)
Krampfanfall	8 (33%)	8 (33%)
Patienten mit mindestens einem UE innerhalb der SOK „Erkrankungen des Nervensystems“ (Tabelle 4-67)	24 (100%)	24 (100%)
Patienten, bei denen mindestens ein UE innerhalb der SMQ Konvulsionen aufgetreten war	23 (96%) 218 UEs	23 (96%) 362 UEs
Häufigste UEs (PT) innerhalb der SMQ Konvulsionen		
Epilepsie	11 (46%) 89 UEs	11 (46%) 175 UEs
Krampfanfall	14 (58%) 110 UEs	14 (58%) 130 UEs
Unerwarteter, rascher Abfall des ML-Scores	0	0
Mindestens ein UE in Zusammenhang mit dem ICV-Zugang (Tabelle 4-68)	7 (29%)	11 (46%) 30 UEs

### *Gesamtzahl der UEs*

Alle Patienten (n=24/24, 100%) entwickelten mindestens ein UE. In Anbetracht der extremen Schwere der Erkrankung und der vulnerablen, pädiatrischen Patientenpopulation, war diese hohe Zahl an UEs nicht unerwartet.

Da nur schwere, schwerwiegende oder anderweitig medizinische bedeutende UEs unter diesen spezifischen Bedingungen relevant sind und die Gesamtzahl auch nicht schwere oder schwerwiegende und nicht patientenrelevante UEs umfasst, wurde die Gesamtzahl der UEs aus rein informativen Gründen dargestellt.

### *Inzidenz von UEs getrennt nach SOK*

Die meisten UEs traten innerhalb der SOKs „Erkrankungen des Nervensystems“ (100%), „Infektionen und parasitäre Erkrankungen“ (100%), „Allgemeine Erkrankungen und

Beschwerden am Verabreichungsort“ (79%), „Erkrankungen des Gastrointestinaltrakt“ (79%), „Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen“ (79%) und „Psychiatrische Erkrankungen“ (50%) auf (Datenschnitt: Juni 2016) [34]. Bei vielen der UEs handelte es sich um typische Krankheitsmanifestationen (insbesondere SOK „Erkrankungen des Nervensystems“) oder um bei Klein- und jungen Kindern zu erwartende Komorbidität. Diese Daten wurden hauptsächlich dargestellt, um üblichen Berichtsstandards zu entsprechen. Die Relevanz dieser Daten wird in den nachfolgenden Kapiteln auf Grundlage der vorab festgelegten SOKs/UEs (PTs) diskutiert, die für die Bewertung von Cerliponase alfa als besonders wichtig eingestuft werden.

#### *Inzidenz sehr häufiger UEs*

Als sehr häufige UEs wurden Fieber (71%), Krampfanfälle (58%), Erbrechen (50%), Infektion der oberen Atemwege (50%), Epilepsie (46%), generalisierte tonisch-klonische Anfälle (46%) und Rhinitis (42%) berichtet (Tabelle 4-61).

Diese Ereignisse sind bei Kleinkindern/Kindern und der Grunderkrankung CLN2 zu erwarten. Da CLN2 eine sehr schwerwiegende und rasch voranschreitende Erkrankung ist, ist die Gesamtzahl der Patienten, die ein UE entwickeln, für die Bewertung des Nutzen- zu Schadenverhältnis von Cerliponase alfa allein nicht ausreichend. Beispielsweise sind leichte Infektionen angesichts des sehr hohen ungedeckten medizinischen Bedarfs für die Patienten und/oder deren Betreuer wahrscheinlich nicht von großer Bedeutung. Diese Daten wurden zur weiteren Information dargestellt. Die Relevanz dieser Daten wird in den nachfolgenden Kapiteln auf Grundlage der vorab festgelegten SOKs/UEs (PTs) diskutiert, die für die Bewertung von Cerliponase alfa als besonders wichtig eingestuft werden.

#### *Inzidenz schwerwiegender UEs und Todesfälle*

Bis Juni 2016 waren insgesamt 51 SUEs bei 19 von 24 Patienten (79%) aufgetreten, davon viele nur einmalig (Tabelle 4-62). Innerhalb der SOK „Infektionen und parasitäre Erkrankungen“ wurden 22 SUEs bei 13 Patienten berichtet, allerdings schätzen die Prüfarzte die Infektionen nur bei zwei Patienten als behandlungsbedingt ein (Datenschnitt Juni 2016). Hinsichtlich der SUEs des Nervensystems (7 SUEs bei 6 Patienten) ist es oftmals schwierig, zwischen Symptomen der Grunderkrankung oder Krankheitsprogression und UEs der Behandlung zu unterscheiden. Dies wird nachfolgend weiter diskutiert.

Todesfälle traten im Verlauf der Studien 190-201/202 nicht auf.

#### *Inzidenz von CTCAE-Grad $\geq 3$ UEs*

CTCAE Grad 3-UEs wurden bis Juni 2016 bei 12/24 Patienten (50%) berichtet und zusätzlich bei einem Patienten ein Grad 4-UE (n=1/24; 4%). Die einzigen UEs Grad  $\geq 3$ , die nach Einschätzung der Prüfarzte mit der Behandlung in Zusammenhang standen, waren Überempfindlichkeitsreaktionen, die bei einer ERT zu erwarten sind.

*UEs von besonderem Interesse: Inzidenz von Überempfindlichkeits- und anaphylaktischen Reaktionen*

Überempfindlichkeits- und anaphylaktische Reaktionen wurden prospektiv unter Einsatz einer breiten MedDRA-Abfrage (SMQ) untersucht. Die Suche nach anaphylaktischen Reaktionen (SMQ) erbrachte kein Ergebnis. Bis Juni 2016 waren bei 14 Patienten (58%) insgesamt 36 Überempfindlichkeitsreaktionen aufgetreten. Folgende UEs (PT) wurden durch die breite SMQ nach Überempfindlichkeitsreaktionen identifiziert: Überempfindlichkeit (n=9, 38%, 15 UEs), jahreszeitbedingte Allergie (n=1, 4%, 2 UEs), Konjunktivitis (n=4, 17%, 5 UEs), Ausschlag (n=2, 8%, 7 AUs), Urtikaria (n=2, 8%, 2 AEs), Dermatitis (n=1, 4%, 1 AE), Kontaktdermatitis (n=2, 8%, 3 UEs) und Stomatitis (n=1, 4%, 1 UE). CTCAE Grad 3 Überempfindlichkeitsreaktionen wurden bei nur drei Patienten (13%) berichtet. Die Daten lieferten keine Hinweise darauf, dass die Zahl der Überempfindlichkeitsreaktionen im Studienverlauf zunimmt (Tabelle 4-65). Bei keinem der Patienten traten mehr als drei Überempfindlichkeitsreaktionen auf (Abbildung 18). Die Überempfindlichkeitsreaktionen manifestierten sich häufig als leichtes Fieber, Erbrechen, Pleozytose oder Reizbarkeit, die nicht den klassisch immunvermittelten Überempfindlichkeitsreaktionen entsprechen und durch die breite SMQ (Überempfindlichkeit) nicht erfasst werden [55].

Neun der insgesamt 15 Überempfindlichkeitsreaktion (PT) wurden als schwerwiegend eingestuft, davon waren allerdings nur drei schwer (Grad 3) und keines sehr schwer oder lebensbedrohlich. Bei keinem der Patienten war eine Dosismodifikation aufgrund des UEs erforderlich, alle UEs waren kontrollierbar und klangen entweder ohne Behandlung oder nach Gabe von Antipyretika, Antihistaminika und/oder Glucokortikoide wieder ab (Anhang 14.3.3 des CSR 190-201). Einige der Patienten, bei denen eine Überempfindlichkeitsreaktion aufgetreten war, waren vor einzelnen Infusion nicht, wie im Studienprotokoll und der Fachinformation empfohlen, mit Antihistaminika in altersgerechter Dosierung vorbehandelt worden.

Wie für einen Wirkstoff biologischen Ursprungs zu erwarten, treten Überempfindlichkeitsreaktionen auch unter Behandlung mit Cerliponase alfa sehr häufig auf. Da alle bislang aufgetretenen Überempfindlichkeitsreaktionen kontrollierbar, keine akut lebensbedrohlich und nur wenige schwer waren und Cerliponase alfa nur unter sorgfältiger Überwachung verabreicht werden darf, werden diese als tolerabel eingeschätzt. In diesem Zusammenhang sind auch die extreme Schwere der Erkrankung und die sehr schlechte Prognose der Patienten zu berücksichtigen.

*UEs von besonderem Interesse: Inzidenz von UEs in zeitlichem Zusammenhang mit den Infusionen („temporally-related events“, TRE)*

UEs, die in zeitlichem Zusammenhang mit den Cerliponase alfa-Infusionen auftraten (TRE), waren prospektiv als UEs von besonderem Interesse definiert. Bis zum Datenschnitt im Juni 2016 hatten alle Patienten (n=24, 100%) mindestens ein TRE entwickelt. Fieber (46%), Überempfindlichkeit (38%), Krampfanfälle (33%), Erbrechen (29%) und Tremor (25%; jeweils PT) wurden am häufigsten berichtet (n=24 Patienten). Bei 7 Patienten (29%) traten CTCAE Grad 3-TRE auf. Krampfanfälle und Tremor sind typische Symptome der Grunderkrankung,

Fieber und Erbrechen können auch als Folge von Begleiterkrankungen, unspezifisch oder im Rahmen von Überempfindlichkeitsreaktionen auftreten. Bei 18 der 24 Studienteilnehmer (75%, 190-201) ergab die Anamnese Epilepsie, bei fünf Patienten (21%) Krampfanfälle und bei zwei Patienten (8%) Tremor.

Unabhängig von der Frage des Kausalzusammenhangs, war keines der TRE akut lebensbedrohlich, die TREs waren mehrheitlich moderat (CTCAE-Grad 2: 67%), und alle waren kontrollierbar.

#### *UEs von besonderem Interesse: Inzidenz von UEs, die das Nervensystem betreffen*

Alle Patienten (n=24/24, 100%) entwickelten mindestens ein UE des Nervensystems, von denen 83% leicht oder mittelschwer waren (CTCAE-Grad 1 oder 2). Durch eine standardisierte MedDRA-Abfrage (SMQ, Konvulsionen) wurden 362 UEs bei 23 Patienten (96%) identifiziert. Die Prüfarzte bewerteten die Mehrzahl dieser Konvulsionsereignisse (340 von 362, 94%) als nicht durch die Behandlung mit Cerliponase alfa bedingt. Die häufigsten Konvulsionsereignisse waren Epilepsie (175 Ereignisse bei 11 Patienten, 46%) und Krampfanfälle (130 Ereignisse bei 7 Patienten, 29%), von denen nach Einschätzung der Prüfarzte nur wenige mit der Behandlung in Zusammenhang standen (4/175 UEs [2.3%] bzw. 14/130 UEs [10.1%]). Insgesamt 59 Konvulsionsereignisse (16,3% von 362 UEs) bei 15 Patienten (63%) traten in zeitlichem Zusammenhang mit Cerliponase alfa auf, d. h. innerhalb von 24 Stunden nach Beginn der Infusion (Tabelle 4-67).

Hinsichtlich der hohen Inzidenz von UEs des Nervensystems und Konvulsionen (SMQ) ist zu berücksichtigen, dass es sich bei nahezu allen dieser UEs um typische Manifestationen der Erkrankung handelt. Bei insgesamt 96% der Studienteilnehmer (n=23/24; 190-201) waren Erkrankungen des Nervensystems bereits aus der Krankenvorgeschichte bekannt (CSR 190-201: Tabelle 14.1.7). Bei 23 der 41 Patienten (56%) der historischen Kontrollgruppe waren Krampfanfälle das erste klinische Zeichen der Erkrankung [56].

Die Inzidenz der berichteten UEs Epilepsie und Krampfanfälle nahm im Studienverlauf tendenziell ab, je länger die Behandlung andauerte. Es ist möglich, dass die Prüfarzte besonders auf UEs des Nervensystems geachtet und diese während der Studie 190-210 und zu Beginn der Studie 190-202 gehäuft berichtet haben („over-reporting“). In diesem Zusammenhang ist zu bedenken, dass es sich um die erste Studie bei Menschen mit einem neuen Wirkstoff biologischen Ursprungs handelt und der Verabreichungsweg ungewöhnlich ist (ICV). Gleichmaßen war die Inzidenz aller UEs, die nach Einschätzung der Prüfarzte mit der Behandlung in Zusammenhang standen, während der ersten 48 Wochen Behandlung (>0-24 Wochen: n=16/24 [67%], >24-48 Wochen: n=17/23 [74%]) höher als nach längerer Behandlung in der Studie 190-202 (>48-72 Wochen: n=10/23 [43%], >72-96 Wochen: n=6/23 [26%]) (Datenschnitt Juni 2016: Tabelle 14.3.1.3.8).

Die hohe Zahl an Konvulsionsereignissen war nicht überraschend. Da nur wenige der Konvulsionsereignisse mit dem Prüfpräparat in Zusammenhang standen (nach Einschätzung der Prüfarzte) oder innerhalb von 24 Stunden nach Beginn der Infusion auftraten, scheint Cerliponase alfa oder der ICV-Zugang diese häufigen Krankheitsmanifestationen nicht negativ

zu beeinflussen oder zu verschlimmern. Dies wird durch die Auswertungen der CLN2-spezifischen QoL-Skala (Elternbericht) bzw. deren Subskala Krampfanfälle gestützt, die keine Zunahme, sondern vielmehr eine im Studienverlauf abnehmende Krampfneigung zeigen (Tabelle 4-56).

*UEs von besonderem Interesse: Unerwartete und rasche Krankheitsprogression (gemessen als Abfall auf der CLN2-spezifischen ML-Skala), die nicht anders erklärbar ist*

Der klinische Review ergab keinen Fall einer unerwartet raschen Krankheitsprogression (gemessen als Abfall des ML-Scores), die nicht anders erklärbar war.

*UEs in Zusammenhang mit dem ICV-Port*

Zum Datenschnitt im Juni 2016, waren bei 11 der 24 Patienten (46%) insgesamt 30 UEs aufgetreten, die mit dem ICV-Port in Zusammenhang standen. Bei 9 Patienten (38%) waren diese UEs milder Ausprägung (CTCAE-Grad 1) und bei zwei Patienten (8%) Grad 3 (Tabelle 4-68). Bei beiden Grad 3-UEs handelte es sich um Infektionen, die einen Austausch des ICV-Ports erforderten. Beide UEs klangen unter Behandlung mit Antibiotika und nach Entfernen des Ports ab, und beide Patienten vertrugen die nachfolgenden Cerliponase alfa-Infusionen.

Nach diesen Ergebnissen treten in Verbindung mit dem ICV-Port keine nicht beherrschbaren Probleme auf.

Während der Studie gab es keine sicherheitsbedingten Therapie- oder Studienabbrüche, und die Lebensqualität (Elternbericht) blieb erhalten oder besserte sich, mit Ausnahme der körperlichen und gesellschaftlichen Funktion (PedsQL™), die sich nur geringfügig verschlechterten. Letzteres könnte Ausdruck der fortschreitenden Erkrankung und der zunehmenden Krankheitslast sein. Es wurden nur relativ wenige UEs berichtet, bei denen ein Kausalzusammenhang mit Cerliponase alfa wahrscheinlich (d. h. plausibel) ist (weitgehend beschränkt auf Überempfindlichkeitsreaktionen), und die UEs waren grundsätzlich kontrollierbar und für die Patienten und/oder deren Familien akzeptabel. Daher sind Cerliponase alfa und der ICV-Port nicht mit bedeutenden oder inakzeptablen Schäden verbunden.

## **Abschließende Beurteilung der Aussagekraft der Nachweise**

### ***Studienqualität***

Zur Bewertung des Zusatznutzens von Cerliponase alfa liegen eine prospektive, einarmige Interventionsstudie (Studie 190-201), deren fortlaufende Folgestudie (Studie 190-202) sowie eine Studie zum natürlichen Krankheitsverlauf vor (Studie 190-901, Original- und aktualisierter Bericht). Grundlage der Wirksamkeitsanalyse (Krankheitsprogression) bildet ein Vergleich zwischen Cerliponase alfa (Studie 190-201/202) und einer historischen Kontrolle aus unbehandelten Patienten, die hinsichtlich wichtiger Prognosefaktoren der Studienpopulation vergleichbar sind („matched“). Nach Kapitel 5, § 5, Absatz (6) der Verfo entspricht die Studie 190-201/202 der Evidenzstufe IV (andere nicht vergleichende Studien).

Gemäß IQWiG (Allgemeine Methoden 4.2) kann auch aus Studien niedrigerer Ergebnis-sicherheit (unterhalb RCT) der Nutzen einer Intervention unter bestimmten Bedingungen

abgeleitet werden: Das neue Arzneimittel kehrt den deterministischen Verlauf einer schwerwiegenden Erkrankung um, der Behandlungseffekt ist „dramatisch“ und der natürliche Verlauf der Erkrankung ist gut dokumentiert [23]. Er wird angemerkt, dass das IQWiG bei Einsatz einer historischen Kontrolle ein „Matching“ nicht explizit fordert [23]. Bei der Bewertung von Ledipasvir/Sofosbuvir hat das IQWiG auf Basis eines nicht adjustierten, historischen Vergleichs mit der Referenztherapie und eines „dramatischen“ Effekts einen Zusatznutzen abgeleitet [67].

Vor Beginn der klinischen Entwicklung von Cerliponase alfa hat BioMarin zahlreiche mögliche Designoptionen geprüft. Randomisierte, kontrollierte Studien repräsentieren den höchsten Grad der evidenzbasierten Medizin. Allerdings war eine aktiv-kontrollierte Studie nicht möglich, da zur Behandlung der CLN2 keine krankheitsmodifizierende Therapie zur Verfügung stand, und eine Placebo-kontrollierte Studie war angesichts der Notwendigkeit, einen ICV-Port zu implantieren, ausgeschlossen. Auch die Möglichkeit einer Parallelgruppe aus unbehandelten Patienten wurde in Betracht gezogen, allerdings wäre diese rasch progredient verlaufende Erkrankung bei den unbehandelten Patienten erwartungsgemäß deutlich und irreversibel vorangeschritten. Der bekannte klinische Verlauf der CLN2 rechtfertigt, zusammen mit dem dringenden, nicht gedeckten medizinischen Bedarf, eine forcierte klinische Entwicklung von Cerliponase alfa [44,68]. Diese Ansicht wird von der wissenschaftlichen Gemeinschaft geteilt (z. B. Augustine et al. 2013 adressiert die Thematik wie folgt: “In a rapidly progressing fatal disease, there is perhaps greater urgency on the part of parents to ensure their child is exposed to an active treatment condition, before the possible window of therapeutic opportunity is lost”) [30]. In CLN2-Säugetiermodellen senkte Cerliponase alfa die Mortalität, was die Erwartung positiver Befunde auch beim Menschen gerechtfertigt erscheinen ließ. Nach bisheriger Kenntnis aus Studien zum natürlichen Verlauf und Auskunft von Experten schreitet die Erkrankung bei den meisten Patienten irreversibel voran, sobald die Patienten in die Phase der Progression („decline phase“) eintreten, von (nahezu) altersgerecht normal hin zu einem fast vegetativen Stadium (das letztlich zum Tod führt). Daher erschien ein Parallelarm aus unbehandelten Patienten schon aus ethischen Gründen fragwürdig. Angesichts der Tatsache, dass die CLN2 eine extrem seltene und extrem schwere Erkrankung ist und die Krankheit, von wenigen Ausnahmen abgesehen, nachgewiesenermaßen vorhersagbar verläuft, wie die Studien zum natürlichen Verlauf belegen, ist die Studie zum historischen Verlauf als Kontrolle geeignet.

Bei Studien mit einer historischen Kontrolle hebt das CHMP die Wichtigkeit eines „Matchings“ zwischen der Studienpopulation und der historischen Kontrolle hervor. Die Erhebungsinstrumente sollten vergleichbar sein und die Endpunkte möglichst gleich [44,68]. BioMarin hat *a priori* (vor den Wirksamkeitsanalysen) verschiedene Algorithmen zur Paarbildung festgelegt, um sicher zu stellen, dass die Studienpopulation und die historische Kontrolle hinsichtlich wichtiger Prognosefaktoren vergleichbar sind. Zusätzlich waren Analysen ohne Paarbildung prospektiv geplant.

Zusammenfassend ist der natürliche Verlauf der CLN2 vorhersagbar, wie dies sowohl die publizierte Literatur als auch die Studie 190-901 zum natürlichen Verlauf übereinstimmend

belegen. Unter diesen spezifischen Bedingungen ist es gerechtfertigt, Studien heranzuziehen, die nicht dem höchsten Evidenzlevel (Ib, RCT) entsprechen.

### *Evidenzstufe*

Entsprechend IQWiG Methoden 4.2 besitzen nicht randomisierte Studien wie z. B. historische kontrollierte Studien grundsätzlich eine sehr geringe qualitative Ergebnissicherheit. Aus einem Prä/Post-Vergleich ohne Kontrollgruppe ist in der Regel kein Beleg für einen Effekt einer Intervention ableitbar. Ausnahmen stellen Erkrankungen mit einem deterministischen (oder quasi-deterministischen) Verlauf dar [23]. Wie oben beschrieben kann die Umkehr eines (quasi)deterministischen Verlaufs durch eine medizinische Intervention ausreichend sein, um den Nutzen eines Arzneimittels zu belegen, vorausgesetzt, der natürliche Verlauf der Erkrankung ist genau vorhersagbar. Ein Effekt wird dann als „dramatisch“ angesehen (d. h. dieser ist durch Störfaktoren allein nicht plausibel erklärbar), wenn er auf einem Niveau von 1% signifikant ist bei einem RR von 5-10 [23,69]. Eine jüngst veröffentlichte GRADE-Richtlinie schlägt Faktoren vor, die die Qualität der Evidenz von insbesondere Beobachtungsstudien erhöhen und ein Heraufstufen der Qualität der Evidenz ermöglichen: Große Effekte (direkte Evidenz, RR 2-5 oder RR 0,5-0,2 ohne plausible Störgrößen); sehr große Effekte mit RR >5 oder RR <0,2 und keine ernsthaften Probleme bezüglich Bias oder unzureichender Präzision (ausreichend enge KIs) [70].

Um das Risiko eines möglichen Bias zu vermindern, wurden in der Studie 190-201 folgende Maßnahmen ergriffen:

- Die Studie 190-201/202 ist eine gut konzipierte, prospektive Interventionsstudie nach heute gültigen Standards für klinische Studien.
- Bei der Studie 190-901 zum natürlichen Verlauf handelt es sich um die bislang größte Längsschnittstudie bei Patienten mit CLN2, und diese repräsentiert die derzeit beste verfügbare Evidenz. Andere historische Längsschnittdaten unter Einsatz der HML-Skala liegen nicht vor. Voraussetzung für eine historische Kontrolle ist, dass die Bewertungsskalen denjenigen der Studie 190-201 vergleichbar sind. Darüberhinaus ist die Studie 190-901 ein von Biomarin unabhängiges EU FP7 Forschungsprojekt und die Daten zum natürlichen Verlauf werden durch Querschnittsdaten des WCMC bestätigt.
- Die Mehrzahl der auswertbaren Patienten der historischen Kontrollgruppe wurde in dem Zentrum in Hamburg rekrutiert (80,5%), welches gleichzeitig auch das größte Zentrum der Studie 190-201/202 ist (50% der eingeschlossenen Patienten). Die Behandlung der CLN2 ist weiterhin rein symptomatisch und es stehen keine krankheitsmodifizierenden Therapien zur Verfügung. Daher ist es unwahrscheinlich, dass sich die Behandlung oder andere Therapiemaßnahmen zwischen den Patienten der historischen Kontrolle (mehrheitlich in dem Zentrum in Hamburg) und den Studienteilnehmern (190-201) erheblich unterscheiden. Sollten dennoch geringfügige Unterschiede zwischen der Studie 190-201/202 und der historischen Kontrolle bestehen, wie z. B. optimierte Behandlung im Rahmen einer prospektiven, klinischen Studie, können diese nicht den deutlichen Effekt von Cerliponase alfa auf die motorischen und sprachlichen Funktionen erklären.



- Zum Vergleich der Studienpopulation (190-201/202) mit der historischen Kontrolle (Studie 190-901) wurden die Haupteignungskriterien der Cerliponase alfa-Studien (190-201/202) eingesetzt (Tabelle 4-31), damit der Datensatz hinsichtlich Alter, der Skalenwerte und der Art der klinischen Bewertungsskalen repräsentativ ist.
- Es liegen verschiedene Sensitivitätsanalysen vor, um zu zeigen, dass das Ergebnis der Primäranalyse robust und über die verschiedenen Analysen konsistent ist. Die Merkmale Krankheitsaktivität (gemessen mittels spezifischer Skalen), Alter sowie Genotyp (jeweils zu Baseline) wurden bei der Paarbildung und zum Vergleich der behandelten und unbehandelten Patienten berücksichtigt. Alle Analysen zum Krankheitsverlauf, gemessen als Abfall auf der ML/HML-Skala, zeigen übereinstimmend einen Vorteil zugunsten Cerliponase alfa gegenüber der historischen Kontrolle, unabhängig von der Auswertungsmethode, der Kriterien zur Paarbildung („Matching“ 1:1 und andere wie „Many-to-one“, um eine möglichst hohe Zahl Paare mit vergleichbaren Baseline-Charakteristika zu erreichen) und des Ausgangspunkts. Der Behandlungseffekt bleibt auch ohne Paarbildung erhalten, wenn alle auswertbaren Patienten der historischen Kontrolle in die Analyse einfließen.
- Die von BioMarin eingereichten Daten zum natürlichen Verlauf liefern verlässliche Nachweise, dass der ML-Score um etwa zwei Punkte pro Jahr abfällt. Der Krankheitsverlauf ist bei der Mehrzahl der Patienten vorhersagbar, sobald die Patienten die progrediente Phase („decline phase“) erreichen, d. h. die Krankheit schreitet irreversibel, nach dem typischen spätinfantilen Phänotyp voran. Unabhängige Arbeitsgruppen bestätigen, dass dies der vorherrschende Phänotyp ist.
- Die ML/HML-Skala bildet klar erkennbare „Meilensteine“ der Krankheitsprogression objektiv ab. Einer der Autoren der Originalpublikation von Steinfeld hat die Performance der ursprünglichen und der modifizierten Version der Skala unabhängig anhand von Videoaufzeichnungen aus der Studie 190-201/202 untersucht; die Analyse war randomisiert und verblindet. Die Performance der HML- und ML-Skala (eingesetzt in der Studie zum natürlichen Verlauf bzw. der Studie 190-201) ist ähnlich. Nach diesen Ergebnissen ist der in der Studie 190-201 beobachtete, deutliche Effekt auf die Krankheitsprogression vermutlich nicht darauf zurückzuführen, dass die modifizierte ML-Skala weniger oder unterschiedlich sensitiv gegenüber Änderungen ist (im Vergleich zur HML-Skala). Die Hälfte der Patienten der Studie 190-201 wurde in dem Zentrum in Hamburg rekrutiert, in dem die HML-Skala entwickelt und bereits viele Jahre vor der Veröffentlichung von Steinfeld et al. im Jahr 2002 [5] eingesetzt wurde.
- Der Effekt von Cerliponase alfa auf die Krankheitsprogression ist so groß („dramatisch“), dass er nicht allein durch unbekannte Störfaktoren oder als Zufallsergebnis erklärbar ist. Darüber hinaus erscheint ein Effekt einer ERT plausibel, vorausgesetzt diese erreicht den Zielort (wofür die Studienergebnisse in geeigneten Tiermodellen sprechen). Die Behandlung richtet sich direkt gegen den zugrunde liegenden pathologischen Prozess.

Angesichts der Tatsache, dass die CLN2 eine sehr seltene und sehr schwerwiegende Erkrankung ist, ist der für Cerliponase alfa eingereichte Datensatz als solide und belastbare Datengrundlage einzustufen. Der Behandlungseffekt von Cerliponase alfa ist über verschiedene Analysen (einschließlich Sensitivitätsanalysen) hinweg konsistent und sehr groß („dramatisch“) im Vergleich zur historischen Kontrolle („matched“), sodass dieser nicht allein durch Störfaktoren erklärbar ist, auch dann nicht, wenn man das hohe Verzerrungspotential einer historischen Kontrolle einbezieht. Dies entspricht den IQWiG Methoden 4.2 [23] und den GRADE-Richtlinien [70].

### **Zusammenfassung - Ausmaß des Zusatznutzens**

Cerliponase alfa ist das erste, zur Behandlung der CLN2 zugelassene Arzneimittel, einer besonders schweren, rasch voranschreitenden und letztlich tödlichen Erkrankung bei jungen Kindern. Der Zusatznutzen von Cerliponase als ein Arzneimittel zur Behandlung eines seltenen Leidens gilt durch die Zulassung als belegt.

Die Studien zum natürlichen Verlauf belegen, dass die Erkrankung vorhersagbar und einem vorbestimmten Muster folgend rasch voranschreitet, sobald die Patienten in die Phase der Progression („decline phase“) eintreten (siehe z. B. Abbildung 9). Letztlich führt die Erkrankung, einer Phase in einem nahezu vegetativen Stadium folgend, zum Tod. Angesichts der Tatsache, dass die krankheitsbedingten neuronalen Schäden irreversibel sind, ist das Behandlungsziel, die durch das Fortschreiten der Erkrankung bedingten Schädigungen und die damit verbundenen Funktionsverluste zu verhindern.

### **Mortalität**

Die Mortalität ist bei dem Patientenkollektiv der Studie 190-201/202 noch nicht bewertbar und entsprechende Daten werden erst nach längerem Follow-up der Patienten vorliegen. Bislang ist keiner der Studienteilnehmer verstorben. Das mittlere Sterbealter für CLN2 bei natürlichem Verlauf liegt bei 10,1 Jahren (siehe Modul 3). Bei Studieneinschluss betrug das mittlere Alter der Patienten in der 190-201 Studie 4,3 (SD 1,3) Jahre. Eine Aussage zur Mortalität kann folglich erst knapp 6 Jahre nach Beginn der Studie getroffen werden, wenn die Patienten ein mittleres Alter von 10,1 Jahren erreicht haben.

### **CLN2-spezifische Bewertungsskalen (ML/HML-Skala)**

Bei fast allen Patienten der Studie 190-201/202 (n=22/23 Patienten der ITT-Population, 97%) war eine 1:1 Paarbildung mit den Patienten der historischen Kontrolle entsprechend der *a priori* festgelegten Kriterien möglich, während etwa die Hälfte der auswertbaren Patienten der historischen Kontrollgruppe aufgrund der Paarbildung aus der Primäranalyse fiel. Die Analyse aller Patienten der historischen Kontrolle (ohne Paarbildung) und die im SAP ursprünglich festgelegte Auswertung (ITT vs. fixe historische Kontrollrate) stützen jedoch die Primäranalyse nach 1:1 Paarbildung und zeigen, dass der Effekt zugunsten von Cerliponase alfa auch nach Ausschluss der ungepaarten Patienten der historischen Kontrolle aus der Primäranalyse vergleichbar groß bleibt. Darüber hinaus zeigen verschiedene Analysen unter Einsatz unterschiedlicher Algorithmen, darunter 1.000 Simulationen zur 1:1 Paarbildung und „Many-

to-one-Matching“, dass die Ergebnisse konsistent sind, unabhängig davon, welche Methode zur Analyse eingesetzt wird.

Das RR für Non-Response (d. h. Abfall des ML-Scores um  $\geq 2$  Punkte/48 Wochen) unter Behandlung mit Cerliponase alfa vs. historische Kontrolle („matched“) betrug 0,16667 ([95% KI: 0,04212; 0,65942],  $p < 0,01$ ) und der entsprechende Kehrwert RR = 6,00. Hinsichtlich der Non-Response auf der Subskala zur Sprache (Abfall um  $\geq 1$  Punkt/48 Wochen), betrug der Unterschied zugunsten von Cerliponase vs. historische Kontrolle RR = 0,125 (Kehrwert: 8) bei engen 95% KIs (0,0327, 0,4773). Analysiert mittels Kaplan-Meier, senkte Cerliponase alfa das Risiko, dass die Erkrankung fortschreitet (d. h. Abfall des ML-Scores um  $\geq 2$  Punkte oder ML-Score = 0), um das Zehnfache gegenüber unbehandelten Patienten der historischen Kontrolle (HR=1,0 [95% KI: 0,03; 0,38],  $p=0,0005$ ). Alle Sensitivitäts- und sonstigen Analysen bestätigen diese Ergebnisse und belegen übereinstimmend den Nutzen von Cerliponase alfa verglichen mit unbehandelten Patienten.

Der Effekt von Cerliponase alfa ist nachhaltig; auch nach einer mindestens 96-wöchigen Behandlung mit Cerliponase alfa ist das Risiko einer Krankheitsprogression (ausgedrückt als Abfall des ML/HML-Scores um  $\geq 2$  Punkte oder ML-Score = 0) stabil um ein erhebliches Ausmaß reduziert, wie die HR von 0,05 ([95% KI 0,01; 0,17],  $p < 0,0001$ ) belegt.

Der Behandlungseffekt von Cerliponase alfa ist gemäß den Vorschlägen des IQWiGs (Methoden 4.2) [23] und der GRADE-Richtlinie [70] als „dramatisch“ zu bewerten und kann durch Störfaktoren nicht plausibel erklärt werden. Cerliponase alfa verzögert den deterministischen Krankheitsverlauf, der ausgehend von einem altersgerecht fast normalen Status über ein Stadium vollständiger Behinderung letztlich zum Tod führt. Die Intervention zeigt positive Behandlungseffekte unabhängig davon, wie stark die Patienten bereits zu Beginn der Behandlung eingeschränkt sind, d. h. über verschiedene ML-Ausgangsscores hinweg. Therapeutisch anzustreben ist trotz alledem eine möglichst frühzeitige Intervention, sodass eine baldigste oder sogar präsymptomatische Diagnose an Bedeutung gewinnen wird [30].

Die Maßnahmen, den Einfluss von Störgrößen möglichst gering zu halten („Matching“ der behandelten und unbehandelten Patienten), und insbesondere der „dramatische“ Effekt von Cerliponase auf die Krankheitsprogression rechtfertigen ein Heraufstufen der Qualität der Evidenz, die üblicherweise mit einer nicht kontrollierten Studie verbunden ist [71].

### **Lebensqualität**

Zur Messung der Lebensqualität wurden der PedsQL™ und ein CLN2-spezifisches QoL-Instrument eingesetzt, das unter Einschluss von betroffenen Eltern entwickelt wurde. Bei beiden Instrumenten handelt es sich nicht um patientenberichtete Instrumente und beide sind für die CLN2-Erkrankung formal noch nicht vollständig validiert. In Anbetracht des jungen Alters der Patienten und der Schwere der Erkrankung ist es nahezu unmöglich, die Lebensqualität direkt auf Basis der Angaben der Patienten zu erfassen. Entsprechend den IQWiG Methoden [23] können bei Interventionen, die Auswirkungen auf indirekt Betroffene wie zum Beispiel Familienangehörige und Pflegepersonen haben, diese berücksichtigt werden. Die Behandlung mit Cerliponase alfa hat keinen negativen Einfluss auf die erhobene

Lebensqualität, sondern führt in den meisten Items zu einer nominellen Verbesserung. In Anbetracht der ohnehin schon sehr großen Belastung der Eltern oder Pflegepersonen, ist es wichtig, dass die Behandlung mit Cerliponase alfa und deren Verabreichungsweg diese Aspekte des täglichen Lebens scheinbar nicht negativ beeinflusst.

### ***Sicherheit und Verträglichkeit***

Die möglicherweise mit Cerliponase alfa assoziierten UEs sind in Anbetracht des extremen Schweregrads der Erkrankung als gut tolerierbar einzustufen. Unter diesen besonderen Bedingungen werden leichte oder mittelschwere UEs bei der Bewertung des Zusatznutzens (oder Schadens) von Cerliponase alfa nicht berücksichtigt. In Anbetracht des „dramatischen“ Effekts von Cerliponase alfa, können nur schwere, schwerwiegende oder lebensbedrohliche UEs oder solche, die zu dauerhaften Beeinträchtigungen führen, das Risiko-Nutzen-Verhältnis ändern. Die Patienten der Studie 190-201 erhielten multiple Begleitmedikationen, die für sich genommen schon mit einer Reihe UEs verbunden sein können und schwer von UEs zu unterscheiden sind, die auf die therapeutische Intervention zurückzuführen sind.

Die Hälfte der Patienten der Studienpopulation (n=12/24, 50%) entwickelte ein unerwünschtes Ereignis vom CTCAE Grad 3. Die einzigen Grad 3-UEs, die nach Einschätzung der Prüfarzte mit der Behandlung in Zusammenhang standen, waren Überempfindlichkeitsreaktionen (PT). Überempfindlichkeitsreaktionen waren auch die häufigsten schwerwiegenden UEs (9 SUEs bei 7/24 Patienten). In Anbetracht der Art des Arzneimittels (ERT) waren Überempfindlichkeitsreaktionen zu erwarten und entsprechend *a priori* als UE von besonderem Interesse definiert. Es trat nur ein CTCAE Grad 4-UE auf (Status epilepticus), welches aus Sicht des Prüfarztes weder mit der Behandlung noch dem ICV-Port in Zusammenhang stand, jedoch ein gut bekanntes Symptom der Grunderkrankung ist. Nach den Ergebnissen scheint Cerliponase alfa die Krampfaktivität und Inzidenz von Konvulsionen nicht zu verändern. Dies sind typische Symptome der Erkrankung und viele Patienten hatten Konvulsionen in ihrer Anamnese.

Unerwünschte Ereignisse, die unter Cerliponase alfa-Behandlung auftraten, sind weitgehend auf Überempfindlichkeitsreaktionen und begleitende Symptome beschränkt; diese sind bei einer ERT zu erwarten. Sicherheitsbedingte Studienabbrüche oder Todesfälle traten nicht auf. Da die schweren, sehr schweren oder schwerwiegenden UEs kontrollierbar waren und ohne Folgeerscheinungen abklangen, stellen diese UEs in Anbetracht der extremen Schwere der Erkrankung keine inakzeptable Toxizität dar. Die Tatsache, dass keiner der Patienten die Behandlung aufgrund eines UEs abgebrochen hat, obwohl alle Patienten mindestens ein UE entwickelt haben, zeigt, dass das Sicherheits- und Verträglichkeitsprofil für die Patienten und/oder deren Eltern/Betreuer insgesamt akzeptabel ist. Dies spiegelt sicherlich zum Teil den hohen, nicht gedeckten medizinischen Bedarf der Patienten wider. Die Verträglichkeit der Behandlung einschließlich der Art der Verabreichung (ICV) wird auch dadurch gestützt, dass sich die Lebensqualität unter Behandlung mit Cerliponase alfa sogar leicht verbesserte. Alle Patienten, die die Studie 190-201 abgeschlossen haben, nehmen weiterhin an der offenen Folgestudie 190-202 teil, was ebenfalls ein Indikator für die gute Akzeptanz von Cerliponase alfa ist.

Da alle schweren, sehr schweren und schwerwiegenden UEs nicht lebensbedrohlich und kontrollierbar waren, ohne Folgeerscheinungen abklagen, und nicht zum Abbruch der Behandlung mit Cerliponase alfa führten, können diese nicht den großen („dramatischen“) Nutzen von Cerliponase alfa hinsichtlich schwerwiegender Manifestationen der Krankheitsprogression aufheben.

In der Gesamtschau ist der dramatische Effekt der Cerliponase alfa-Behandlung in der Endpunktkategorie Morbidität unter Berücksichtigung der anderen Endpunktkategorien als eine langfristige Freiheit von schwerwiegenden Symptomen einzustufen, die sich bei unbehandelten Patienten im Laufe der Krankheitsprogression manifestieren. Diese nachhaltige und bisher nicht erreichte große Verbesserung des therapielevanten Nutzens ist entsprechend § 5 (7) AM NutzenV als **erheblich** einzustufen.

#### 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 4-74: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
Patienten mit neuronaler Ceroid-Lipofuszinose Typ 2 (CLN2, Tripeptidyl-Peptidase-1-(TPP 1)-Mangel)	Erheblich

#### 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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

##### 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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **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>13</sup>, Molenberghs 2010<sup>14</sup>). Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrunde liegende Anwendungsgebiet und das zu bewertende Arzneimittel sowie die Vergleichstherapie erlauben.

Eine Möglichkeit der Verwendung von Surrogatendpunkten ohne abschließende Validierung stellt die Anwendung des Konzepts eines sogenannten Surrogate-Threshold-Effekts (STE) (Burzykowski 2006<sup>15</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>16</sup>) soll ausreichend begründet

<sup>13</sup> Burzykowski T (Ed.): *The evaluation of surrogate endpoints*. New York: Springer; 2005.

<sup>14</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>15</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>16</sup> Weir CJ, Walley RJ: *Statistical evaluation of biomarkers as surrogate endpoints: a literature review*. *Stat Med* 2006; 25(2): 183-203.

werden, insbesondere dann, wenn als Datengrundlage nur eine einzige Studie verwendet werden soll.

Berichten Sie zu den Studien zur Validierung oder zur Begründung für die Verwendung von Surrogatendpunkten mindestens folgende Informationen:

- Patientenpopulation
- Intervention
- Kontrolle
- Datenherkunft
- verwendete Methodik
- entsprechende Ergebnisse (abhängig von der Methode)
- Untersuchungen zur Robustheit
- ggf. Untersuchungen zur Übertragbarkeit

*Sofern Sie im Dossier Ergebnisse zu Surrogatendpunkten eingereicht haben, benennen Sie die Gründe für die Verwendung von Surrogatendpunkten. Beschreiben Sie, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.*

Für Arzneimittel zur Behandlung seltener Leiden sind nur die Abschnitte 4.4.2 und 4.4.3 auszufüllen [18].

#### **4.6 Liste der eingeschlossenen Studien**

*Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).*

Studie	Datenquelle
190-201	Finaler Studienbericht 190-201 (27.04.2016) [26] Studienprotokoll 190-201 Amendment 5 (24.09.2014) [37] Finaler SAP v3 190-201 (25.03.2016) [51]
190-202	Studienbericht 190-202 (09.05.2016) [27] Studienbericht 190-202 Addendum (03.05.2016) [66] Studienprotokoll 190-202 A-03 (Amendment 3, 26.02.2016) [45] FDA Response (20 Jul 2016) [34]
190-901 (Natural history report)	Studie 190-901, Original-Analyse (30.06.2015) [39] Studie 190-901, Supplemental Report (01.05.2016) [56] EVA-13748-01, Clinician-Reported Outcome Evidence Dossier (18.04.2016) [24]
ISE, Integrated Summary of Efficacy (historische Vergleiche)	Integrated Summary of Efficacy (M2.7.3 Summary of Clinical Efficacy) [57] SAP der „Integrated Summary of Efficacy“ (ISE SAP; 25.04.2016) [58]  <i>96-Wochen Follow-up:</i> FDA Response (01 Feb 2017) [59] SAP ISE Auswertung der 96-Wochen-Daten [60]

#### 4.7 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. Nickel M, Jacoby D, Lezius S, Down M, Simonati A, Genter F, u. a. Natural History of CLN2 Disease: Quantitative Assessment of Disease Characteristics and Rate of Progression. *Neuropediatrics*. April 2016;47(S 01):FV04-03–03.
2. Williams RE, Aberg L, Autti T, Goebel HH, Kohlschütter A, Lönnqvist T. Diagnosis of the neuronal ceroid lipofuscinoses: an update. *Biochim Biophys Acta*. Oktober 2006;1762(10):865–72.
3. Chang M, Cooper JD, Davidson BL, van Diggelen OP, Elleder M, Goebel HH, u. a. CLN2. In: Mole S, Williams R, and Goebel HH, Hrsg *The neuronal ceroid lipofuscinoses (Batten Disease)*. 2. Aufl. Oxford, UK: Oxford University Press; 2011. S. 80–109.
4. Pérez-Poyato MS, Marfa MP, Abizanda IF, Rodriguez-Revenga L, Sánchez VC, González MJM, u. a. Late infantile neuronal ceroid lipofuscinosis: mutations in the CLN2 gene and clinical course in Spanish patients. *J Child Neurol*. April 2013;28(4):470–8.
5. Steinfeld R, Heim P, von Gregory H, Meyer K, Ullrich K, Goebel HH, u. a. Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations. *Am J Med Genet*. 1. November 2002;112(4):347–54.



6. Worgall S, Kekatpure MV, Heier L, Ballon D, Dyke JP, Shungu D, u. a. Neurological deterioration in late infantile neuronal ceroid lipofuscinosis. *Neurology*. 7. August 2007;69(6):521–35.
7. Moore SJ, Buckley DJ, MacMillan A, Marshall HD, Steele L, Ray PN, u. a. The clinical and genetic epidemiology of neuronal ceroid lipofuscinosis in Newfoundland. *Clin Genet*. September 2008;74(3):213–22.
8. Sleat DE, Gin RM, Sohar I, Wisniewski K, Sklower-Brooks S, Pullarkat RK, u. a. Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. *Am J Hum Genet*. Juni 1999;64(6):1511–23.
9. Fietz M, AlSayed M, Burke D, Cohen-Pfeffer J, Cooper JD, Dvořáková L, u. a. Diagnosis of neuronal ceroid lipofuscinosis type 2 (CLN2 disease): Expert recommendations for early detection and laboratory diagnosis. *Mol Genet Metab*. September 2016;119(1–2):160–7.
10. Schulz A, Kohlschütter A, Mink J, Simonati A, Williams R. NCL diseases — clinical perspectives. *Biochim Biophys Acta*. November 2013;1832(11):1801–6.
11. Williams RE, Mole SE. New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses. *Neurology*. 10. Juli 2012;79(2):183–91.
12. Geraets RD, Koh S yon, Hastings ML, Kielian T, Pearce DA, Weimer JM. Moving towards effective therapeutic strategies for Neuronal Ceroid Lipofuscinosis. *Orphanet J Rare Dis*. 16. April 2016;11:40.
13. Banks WA. Characteristics of compounds that cross the blood-brain barrier. *BMC Neurol*. 12. Juni 2009;9 Suppl 1:S3.
14. European Medicines Agency. Committee for Orphan Medicinal Products. Public summary of opinion on orphan designation. Recombinant human tripeptidyl-peptidase 1 for the treatment of neuronal ceroid lipofuscinosis type 2. EMA/COMP/92274/2013 [Internet]. 2013 [zitiert 9. November 2016]. Verfügbar unter: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Orphan\\_designation/2013/04/WC500141555.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Orphan_designation/2013/04/WC500141555.pdf)
15. Sozialgesetzbuch (SGB V) Fünftes Buch. Gesetzliche Krankenversicherung. Zuletzt geändert durch Art. 2a G v. 18.7.2016 I 1710 [Internet]. 2016 [zitiert 9. November 2016]. Verfügbar unter: <http://www.sozialgesetzbuch-sgb.de/sgbv/35a.html>
16. Gemeinsamer Bundesausschuss (G-BA). Verfahrensordnung des Gemeinsamen Bundesausschusses in der Fassung vom 18. Dezember 2008 veröffentlicht im Bundesanzeiger Nr. 84a (Beilage) vom 10. Juni 2009 in Kraft getreten am 1. April 2009, zuletzt geändert am 20. Oktober 2016 veröffentlicht im Bundesanzeiger BAnz AT 24.02.2017 B1, in Kraft getreten am 20. Januar 2017. 20. Januar 2017 [zitiert 9. November 2016]; Verfügbar unter: [https://www.g-ba.de/downloads/62-492-1331/VerfO\\_2016-10-20\\_iK-2017-01-20.pdf](https://www.g-ba.de/downloads/62-492-1331/VerfO_2016-10-20_iK-2017-01-20.pdf)

17. Project Final Report. A Treatment-Oriented Research Project of NCL Disorder as a Major Cause of Dementia in Childhood. DEM-CHILD; 2014 Sep.
18. Gemeinsamer Bundesausschuss (G-BA). Erstellung und Einreichung eines Dossiers zur Nutzenbewertung gemäß § 35a SGB V. Format und Gliederung des Dossiers, einzureichende Unterlagen, Vorgaben für technische Standards [Internet]. 2013 [zitiert 9. November 2016]. Verfügbar unter: [https://www.g-ba.de/downloads/17-98-3518/2013-04-18\\_An12\\_1\\_Erstellung-Einreichung-Dossier.pdf](https://www.g-ba.de/downloads/17-98-3518/2013-04-18_An12_1_Erstellung-Einreichung-Dossier.pdf)
19. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. März 2004;94(3):361–6.
20. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, u. a. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 16. Oktober 2007;147(8):573–7.
21. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, u. a. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 20. Oktober 2007;335(7624):806–8.
22. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, u. a. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. Oktober 2015;12(10):e1001885.
23. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. Version 4.2 vom 22.04.2015 [Internet]. 2015 [zitiert 16. November 2016]. Verfügbar unter: [https://www.iqwig.de/download/IQWiG\\_Methoden\\_Version\\_4-2.pdf](https://www.iqwig.de/download/IQWiG_Methoden_Version_4-2.pdf)
24. Wyrwich K. EVA-13748-01. The 0-6 Hamburg Scale (HML and ML) in Patients with Neuronal Ceroid Lipofuscinosis Type 2 (CLN2). BioMarin Studies 190-901 (HML) and 190-201(ML). A Clinician-Reported Outcome Evidence Dossier Prepared For: BioMarin. Evidera; 2016 Mai.
25. Löbel U, Sedlacik J, Nickel M, Lezius S, Fiehler J, Nestrasil I, u. a. Volumetric Description of Brain Atrophy in Neuronal Ceroid Lipofuscinosis 2: Supratentorial Gray Matter Shows Uniform Disease Progression. *AJNR Am J Neuroradiol*. 26. Mai 2016;
26. Noonberg S, Jacoby D, Slasor P, Nickas J, Malandro L, Escandon R, u. a. A Phase 1/2 Open-Label Dose-Escalation Study to Evaluate Safety, Tolerability, Pharmacokinetics, and Efficacy of Intracerebroventricular BMN 190 in Patients with Late-Infantile Neuronal Ceroid Lipofuscinosis (CLN2) Disease. Clinical Study Report 190-201. 27. April 2016;
27. Noonberg S, Jacoby D, Slasor P, Nickas J, Malandro L, Escandon R, u. a. A Multicenter, Multinational, Extension Study to Evaluate the Long-Term Efficacy and Safety of BMN 190 in Patients with CLN2 Disease. Report 190-202. Mai 2016;
28. EMA. CHMP assessment report. Brineura. 2017.

29. Kousi M, Lehesjoki A-E, Mole SE. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. *Hum Mutat.* Januar 2012;33(1):42–63.
30. Augustine EF, Adams HR, Mink JW. Clinical trials in rare disease: challenges and opportunities. *J Child Neurol.* September 2013;28(9):1142–50.
31. Hobert JA, Dawson G. Neuronal ceroid lipofuscinoses therapeutic strategies: past, present and future. *Biochim Biophys Acta.* Oktober 2006;1762(10):945–53.
32. Bundesministeriums der Justiz und für Verbraucherschutz. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV), Zuletzt geändert durch Art. 2 G v. 27.3.2014 I 261 [Internet]. 2014 [zitiert 10. November 2016]. Verfügbar unter: <https://www.gesetze-im-internet.de/bundesrecht/am-nutzenv/gesamt.pdf>
33. BioMarin Pharmaceutical Inc. A Phase 2 Open-Label Study to Evaluate Safety, Tolerability, and Efficacy of Intracerebroventricular BMN 190 in Patients with CLN2 Disease. EudraCT Number: 2015-000891-85 [Internet]. 2015 [zitiert 24. Dezember 2016]. Verfügbar unter: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2015-000891-85/DE>
34. BioMarin Pharmaceutical Inc. Study 190-202. RESPONSE TO REQUEST FOR INFORMATION DATED JULY 20, 2016. 2016.
35. Worgall S, Sondhi D, Hackett NR, Kosofsky B, Kekatpure MV, Neyzi N, u. a. Treatment of late infantile neuronal ceroid lipofuscinosis by CNS administration of a serotype 2 adeno-associated virus expressing CLN2 cDNA. *Hum Gene Ther.* Mai 2008;19(5):463–74.
36. Dyke JP, Sondhi D, Voss HU, Shungu DC, Mao X, Yohay K, u. a. Assessment of disease severity in late infantile neuronal ceroid lipofuscinosis using multiparametric MR imaging. *AJNR Am J Neuroradiol.* April 2013;34(4):884–9.
37. BioMarin Pharmaceutical Inc. Clinical Study Protocol. A Phase 1/2 Open-Label Dose-Escalation Study to Evaluate Safety, Tolerability, Pharmacokinetics, and Efficacy of Intracerebroventricular BMN 190 in Patients with Late-Infantile Neuronal Ceroid Lipofuscinosis (CLN2) Disease. Amendment 5. 2014 Sep.
38. Crystal RG, Sondhi D, Hackett NR, Kaminsky SM, Worgall S, Stieg P, u. a. Clinical protocol. Administration of a replication-deficient adeno-associated virus gene transfer vector expressing the human CLN2 cDNA to the brain of children with late infantile neuronal ceroid lipofuscinosis. *Hum Gene Ther.* November 2004;15(11):1131–54.
39. BioMarin Pharmaceutical Inc., Jacoby D, Fong M, Slasor P. Natural History of Late-Infantile CLN2 Disease: Quantitative Assessment of Disease Characteristics, Rate of Progression, and Magnetic Resonance Imaging Findings. Natural History Report No. 190-901. 30. Juni 2015;

40. Specchio N, Williams RE, Adams HR, Blohm M, Cohen-Pfeffer JL, de los Reyes E, u. a. Expert Opinion on the Management of CLN2 Disease. Society for the Study of Inborn Errors of Metabolism (SSIEM) Annual Symposium: 6–9 September, 2016. In Italy: Rome; 2016.
41. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* Oktober 1968;70(4):213–20.
42. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics.* Juni 1977;33(2):363–74.
43. Varni JW, Sherman SA, Burwinkle TM, Dickinson PE, Dixon P. The PedsQL Family Impact Module: preliminary reliability and validity. *Health Qual Life Outcomes.* 27. September 2004;2:55.
44. European Medicines Agency. Scientific advice. BMN 190. EMA/CHMP/SAWP/549581/2013. Procedure No. EMEA/H/SA/2544/1/2013/PED/III. 19. September 2013;
45. BioMarin Pharmaceutical Inc. A Multicenter, Multinational, Extension Study to Evaluate the Long-Term Efficacy and Safety of BMN 190 in Patients with CLN2 Disease. Protocol 190-202 Amendment 3. 2016 Feb.
46. European Medicines Agency. Guideline on clinical trials in small populations. CHMP/EWP/83561/2005 [Internet]. 2006 [zitiert 11. Dezember 2016]. Verfügbar unter: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003615.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf)
47. Matza LS, Patrick DL, Riley AW, Alexander JJ, Rajmil L, Pleil AM, u. a. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* Juni 2013;16(4):461–79.
48. Msall ME. Measuring functional skills in preschool children at risk for neurodevelopmental disabilities. *Ment Retard Dev Disabil Res Rev.* 2005;11(3):263–73.
49. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr Off J Ambul Pediatr Assoc.* Dezember 2003;3(6):329–41.
50. Varni, James W. Scaling and scoring of the Pediatric Quality of Life Inventory PedsQL. Version 14 [Internet]. 2014 [zitiert 19. November 2016]. Verfügbar unter: <http://www.pedsqol.org/PedsQL-Scoring.pdf>
51. BioMarin Pharmaceutical Inc., Bray L, Slasor P, Jacoby D, Veras D, Cramm T, u. a. Final Statistical Analysis Plan. Study BMN 190-201. A Phase 1/2 Open-Label Dose-Escalation Study to Evaluate Safety, Tolerability, Pharmacokinetics, and Efficacy of Intracerebroventricular BMN 190 in Patients with Late-infantile Neuronal Ceroid Lipofuscinosis (CLN2) Disease. Version 3.0. 2016 März.

52. Bullinger M, Blome C, Sommer R, Lohrberg D, Augustin M. [Health-related quality of life: a pivotal endpoint in benefit assessment of medical procedures]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. März 2015;58(3):283–90.
53. Ballinger, Rachel, Eliasson, Lina, West, Andrea, Schulz, Angela, Peasgood, Tessa, Jain, Mohit, u. a. The burden of CLN2 disease on families: home-based surveys with caregivers in Germany and the United Kingdom. O42. In Boston, USA; 2016 [zitiert 12. Dezember 2016]. S. 52. Verfügbar unter: [http://nclboston2016.com/wp-content/uploads/2016/07/NCL-2016-Boston\\_Program\\_final.pdf](http://nclboston2016.com/wp-content/uploads/2016/07/NCL-2016-Boston_Program_final.pdf)
54. Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, u. a. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 16. November 2004;141(10):781–8.
55. BioMarin Pharmaceutical Inc. Module 2.7.4 Summary of Clinical Safety. 2016 Mai.
56. BioMarin Pharmaceutical Inc., Jacoby D, Fong M, Slasor P. Updated Analysis of the Natural History of CLN2 Disease: Estimated Rate of Decline from the DEM-CHILD Multi-Center Clinical NCL Database. 190-901 Supplemental Report. 2016 Mai.
57. BioMarin Pharmaceutical Inc. Module 2.7.3. Summary of Clinical Efficacy. 2016 Mai.
58. BioMarin Pharmaceutical Inc. Final Statistical Analysis Plan. BMN 190 Integrated Summary of Efficacy. Version 3. 2016 März.
59. BioMarin Pharmaceutical Inc. RESPONSE TO REQUEST FOR INFORMATION FROM FEBRUARY 1, 2017. 2017.
60. BioMarin Pharmaceutical Inc. Supplemental Statistical Analysis Plan For the Analysis of Week 96 Efficacy Data (cerliponase alfa). 2017 März.
61. BioMarin Pharmaceutical Inc. STATISTICAL ANALYSES, DATASETS, AND PROGRAMS PER THE REVISED SUPPLEMENTAL SAP (VERSION DATED 20 MARCH 2017) (PART 1 OF 2). 2017 März.
62. BioMarin Pharmaceutical Inc. STATISTICAL ANALYSES, DATASETS, AND PROGRAMS PER THE REVISED SUPPLEMENTAL SAP (VERSION DATED 20 MARCH 2017) (PART 2 OF 2). 2017 März.
63. Santorelli FM, Garavaglia B, Cardona F, Nardocci N, Bernardina BD, Sartori S, u. a. Molecular epidemiology of childhood neuronal ceroid-lipofuscinosis in Italy. Orphanet J Rare Dis. 2. Februar 2013;8:19.
64. Cochrane Deutschland, AWMF. Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung [Internet]. 2016 [zitiert 22. November 2016]. Verfügbar unter: <http://www.cochrane.de/de/news/bewertung-des-biasrisikos-risiko-systematischer-fehler-klinischen-studien>

65. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. Mai 2003;41(5):582–92.
66. BioMarin Pharmaceutical Inc. 190-202 Addendum. 2016 Juli.
67. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. IQWiG-Berichte – Nr. 282. Ledipasvir/Sofosbuvir – Nutzenbewertung gemäß § 35a SGB V [Internet]. 2015 [zitiert 7. Dezember 2016]. Verfügbar unter: [https://www.iqwig.de/download/A14-44\\_Ledipasvir-Sofosbuvir\\_Nutzenbewertung\\_35a-SGB-V.pdf](https://www.iqwig.de/download/A14-44_Ledipasvir-Sofosbuvir_Nutzenbewertung_35a-SGB-V.pdf)
68. European Medicines Agency. Follow-up Protocol Assistance BMN 190. EMEA/CHMP/SAWP/209064/2007, Procedure No.: EMEA/H/SA/2544/1/FU/1/2014/PA/PED/II. 24. Juli 2014;
69. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 17. Februar 2007;334(7589):349–51.
70. Kien C, Gartlehner G, Kaminski-Hartenthaler A, Meerpohl JJ, Flamm M, Langer G, u. a. [GRADE guidelines: 9. Rating up the quality of evidence]. *Z Evidenz Fortbild Qual Im Gesundheitswesen*. 2013;107(3):249–55.
71. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, u. a. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. Dezember 2011;64(12):1311–6.
72. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, u. a. [The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting of observational studies]. *Internist*. Juni 2008;49(6):688–93.
73. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, u. a. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 16. Oktober 2007;4(10):e297.
74. Katz ML, Coates JR, Sibigtroth CM, Taylor JD, Carpentier M, Young WM, u. a. Enzyme replacement therapy attenuates disease progression in a canine model of late-infantile neuronal ceroid lipofuscinosis (CLN2 disease). *J Neurosci Res*. November 2014;92(11):1591–8.

**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>	08.12.2010	
<b>Zeitsegment</b>	1980 to 2010 week 50	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle <sup>17</sup> ] – Strategy minimizing difference between sensitivity and specificity	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	Meglitinide/	848
2	Nateglinide/	1686
3	Repaglinide/	2118
4	(glinid* or meglitinid* or nateglinid* or repaglinid*).ab,ti.	1069
5	(starlix or novonorm or novo norm or prandin).ab,ti.	32
6	(105816-04-4 or 135062-02-1).rn.	2854
7	or/1-6	3467
8	Diabetes mellitus/	224164
9	Non Insulin dependent Diabetes mellitus/	91081
10	(diabet* or niddm or t2dm).ab,ti.	379777
11	or/8-10	454517
12	(random* or double-blind*).tw.	650136
13	placebo*.mp.	243550
14	or/12-13	773621
15	and/7,11,14	719

<sup>17</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**

Nicht zutreffend.

**Anhang 4-A2: Suche nach RCT für indirekte Vergleiche**

Nicht zutreffend.

**Anhang 4-A3: Suche nach nicht randomisierten vergleichenden Studien**

Nicht zutreffend.

**Anhang 4-A4: Suche nach weiteren Untersuchungen**

Nicht zutreffend.



**Anhang 4-B: Suchstrategien – Suche in Studienregistern**

Geben Sie nachfolgend die Suchstrategien für die Suche(n) in Studienregistern an. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Für jedes durchsuchte Studienregister ist eine separate Strategie darzustellen. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters (z. B. [clinicaltrials.gov](http://www.clinicaltrials.gov)), die Internetadresse, unter der das Studienregister erreichbar ist (z. B. <http://www.clinicaltrials.gov>), das Datum der Suche, die verwendete Suchstrategie und die resultierenden Treffer an. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

<b>Studienregister</b>	clinicaltrials.gov
<b>Internetadresse</b>	<a href="http://www.clinicaltrials.gov">http://www.clinicaltrials.gov</a>
<b>Datum der Suche</b>	08.12.2010
<b>Suchstrategie</b>	(Starlix OR Novonorm OR Prandin OR Nateglinid OR Repaglinid) [ALL-FIELDS] AND ("Phase II" OR "Phase III" OR "Phase IV") [PHASE]
<b>Treffer</b>	23

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

Nicht zutreffend.

**Anhang 4-B2: Suche nach RCT für indirekte Vergleiche**

Nicht zutreffend.

**Anhang 4-B3: Suche nach nicht randomisierten vergleichenden Studien**

Nicht zutreffend.

**Anhang 4-B4: Suche nach weiteren Untersuchungen**

Nicht zutreffend.

**Anhang 4-C: Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche)**

*Listen Sie nachfolgend die im Volltext gesichteten und ausgeschlossenen Dokumente aus der / den bibliografischen Literaturrecherche(n) auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.*

**Anhang 4-C1: Suche nach RCT mit dem zu bewertenden Arzneimittel**

Nicht zutreffend.

**Anhang 4-C2: Suche nach RCT für indirekte Vergleiche**

Nicht zutreffend.

**Anhang 4-C3: Suche nach nicht randomisierten vergleichenden Studien**

Nicht zutreffend.

**Anhang 4-C4: Suche nach weiteren Untersuchungen**

Nicht zutreffend.

**Anhang 4-D: Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern)**

*Listen Sie nachfolgend die durch die Studienregistersuche(n) identifizierten, aber ausgeschlossenen Studien auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.*

**Anhang 4-D1: Suche nach RCT mit dem zu bewertenden Arzneimittel**

Nicht zutreffend.

**Anhang 4-D2: Suche nach RCT für indirekte Vergleiche**

Nicht zutreffend.

**Anhang 4-D3: Suche nach nicht randomisierten vergleichenden Studien**

Nicht zutreffend.

**Anhang 4-D4: Suche nach weiteren Untersuchungen**

Nicht zutreffend.

**Anhang 4-E: Methodik der eingeschlossenen Studien**

*Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.4 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-75 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 4-75 inklusive eines Flow-Charts für den Patientenfluss.*

Die nachfolgende Tabelle 4-75 in Anhang 4-E wurde nach TREND („Transparent Reporting of Evaluations with Nonrandomized Designs“) 2004 bearbeitet [19]. Die Anwendung des TREND-Statements zur Beschreibung des Studiendesigns und der Studienmethodik für Studie 190-201 ist angezeigt, da zur Bewertung des Zusatznutzens von Cerliponase alfa keine randomisierten Studien vorliegen. Im finalen Studienbericht wurde der primäre Endpunkt entsprechend des SAPs Version 3 ausgewertet. Auf Empfehlung der Zulassungsbehörden wurde der SAP für die ISE („Integrated Summary of Efficacy“) ergänzt und die primäre Responder-Analyse dahingehend modifiziert, dass die Wirksamkeit von Cerliponase alfa mit einer historischen Kontrolle nach Paarbildung (matching) verglichen wurde (in der Tabelle gekennzeichnet durch „SAP ISE“). Aus Gründen der Transparenz sind in der folgenden Tabelle beide Analysen dargestellt.

Die nachfolgende Tabelle 4-77, die das Design und die Methodik der nicht interventionellen Studie 190-901 zum natürlichen Verlauf der Erkrankung beschreibt, wurde nach STROBE (“The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies“) 2007/2008 bearbeitet [20,72,73]. Die Anwendung des STROBE-Statements zur Beschreibung des Studiendesigns und der Studienmethodik erscheint angezeigt, da es sich um eine Beobachtungsstudie handelt. Das RECORD-Statement („REporting of studies Conducted using Observational Routinely-collected health Data“) wurde bei der Bearbeitung berücksichtigt [22]. Es wird angemerkt, dass eine Darstellung dieser nicht interventionellen Studie zum natürlichen Verlauf der Erkrankung nach STROBE nicht erforderlich gewesen wäre. Dies wurde dennoch als sinnvoll erachtet, um eine möglichst transparente Darstellung z.B. der Patientenselektion der historischen Kontrolle zu gewährleisten. Nach IQWiG Methoden 4.2 [23] kann unter bestimmten Bedingungen aus einer gut dokumentierte Fallserie zum natürlichen Verlauf ein Nutzen einer Intervention abgeleitet werden. Daher erscheint eine möglichst detaillierte Darstellung der Studie zum natürlichen Verlauf bei unbehandelten CLN2-Patienten sinnvoll.

Tabelle 4-75 (Anhang): Studiendesign und -methodik für Studie 190-201

Item <sup>a</sup>	Charakteristikum	Studieninformation
-	<b>Studienziel</b>	
2	Hintergrund / Rationale <ul style="list-style-type: none"> <li>• Wissenschaftlicher Hintergrund und Erläuterung der Rationale</li> <li>• Theorien bei der Konzeption von Verhaltensinterventionen (nicht zutreffend)</li> </ul>	<p><b>Wissenschaftlicher Hintergrund und Erläuterung der Rationale</b></p> <p>Bei der neuronalen Ceroid-Lipofucinose Typ 2 (CLN2) handelt es sich um eine sehr seltene, autosomal rezessiv vererbte neurodegenerative Erkrankung, die zu den lysosomalen Speicherkrankheiten zählt. Die Erkrankung wird durch Mutationen des <i>TPP1/CLN2</i>-Gens verursacht, die zu einer verminderten Aktivität der lysosomalen Exopeptidase TPP1 führen [9]. Wie bei anderen NCL kommt es zu einer Akkumulation autofluoreszenten Speichermaterials innerhalb der Lysosomen von Neuronen und anderen Zelltypen verbunden mit neuronalem Zelluntergang [9,10].</p> <p>Der klassische Phänotyp der CLN2 ist durch Krampfanfälle und/oder Ataxie in der spätinfantilen Phase (etwa im Alter von 2-4 Jahren) oft in Kombination mit einer vorausgegangenen, verzögerten Sprachentwicklung gekennzeichnet. Bei Patienten mit atypischem Verlauf tritt die Erkrankung später auf und in einigen Fällen ist die Lebenserwartung etwas höher [9]. Ab einem Alter von etwa fünf Jahren sind die Kinder in der Regel vollständig abhängig [3,5]. Krankheitsmodifizierende oder kurative Therapien existieren bislang für keine der NCL [10].</p> <p>Zusammenfassend ist die CLN2 eine sehr schwerwiegende, rasch fortschreitende neurodegenerative Erkrankung im Frühkindalter, die letztlich tödlich verläuft. Daher werden dringend effektive Behandlungen benötigt, um das Fortschreiten der Erkrankung sowie den Verlust entscheidender kindlicher Entwicklungen so früh und so lange wie möglich zu verhindern.</p> <p>Cerliponase alfa (Brineura®) ist eine rekombinante Form der humanen TPP1, des bei Patienten mit CLN2 fehlerhaften bzw. fehlenden Enzyms. Die Enzyersatztherapie mit Cerliponase alfa hat zum Ziel, die Enzymaktivität im Zielgewebe wiederherzustellen. Die intrathekale Gabe von Cerliponase alfa verbesserte in zwei etablierten Tiermodellen der CLN2 signifikant den klinischen Verlauf, insbesondere wenn die Tiere bereits kurz nach der Geburt behandelt wurden. Der erhebliche therapeutische Bedarf sowie die</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		vielversprechenden Ergebnisse der tierexperimentellen Studien rechtfertigen eine klinische Studie mit Cerliponase alfa (ICV) bei Kleinkindern/Kindern mit CLN2.
-	<b>Methoden</b>	
3	<p>Studienteilnehmer</p> <ul style="list-style-type: none"> <li>• Eignungskriterien für Studienteilnehmer, einschließlich der Kriterien auf verschiedenen Ebenen des Rekrutierungsplans / Plan zur Stichprobennahme (sampling plan) (z. B. Städte, Kliniken, Patienten)</li> <li>• Rekrutierungsmethode (z. B. Überweisung, Selbstauswahl) einschließlich der Methode zur Stichprobennahme (sampling method), falls ein systematischer Plan zur Stichprobennahme verwendet wurde</li> <li>• Vorgehensweise bei der Rekrutierung</li> <li>• Setting der Datenerhebung und Orte, an denen die Daten erhoben wurden</li> </ul>	<p><b>Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>- Gesicherte CLN2-Diagnose (TPP1-Enzymaktivität im Blut und CLN2-Genanalyse)</li> <li>- Leichte bis mittelschwere Erkrankung (Score von 3-6 auf der modifizierten ML Hamburg-Skala und Score von mindestens 1 in jeder der beiden Domänen)</li> <li>- Schriftliche Einverständniserklärung durch ein Elternteil oder gesetzlichen Vertreter und falls möglich Zustimmung des Patienten</li> <li>- Fähigkeit, an den protokollspezifischen Untersuchungen teilzunehmen</li> <li>- Krampfanfälle und begleitende Anti-Epileptika aus Sicht des Prüfarztes stabil</li> </ul> <p><b>Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>- Alter &lt;3 Jahre oder ≥16 Jahre zum Zeitpunkt des Studieneinschlusses</li> <li>- Andere, neurologische Erbkrankheit wie z.B. andere CLN oder nicht mit CLN2 in Zusammenhang stehende Krampfanfälle (Einschluss von Patienten mit Fieberkrämpfen möglich)</li> <li>- Andere neurologische Erkrankung vor Studieneinschluss, die zum Verlust kognitiver Fähigkeiten geführt haben könnte (z. B. Trauma, Meningitis, Blutung)</li> <li>- Unterstützende Beatmaßungsmaßnahmen ausgenommen nicht invasive Maßnahmen in der Nacht</li> <li>- Vorausgegangene Stammzell- oder Gentherapie oder ERT zur Behandlung der CLN2</li> <li>- Kontraindikationen für einen neurochirurgischen Eingriff (z.B. angeborener Herzfehler, schwere respiratorische Insuffizienz oder Gerinnungsstörung)</li> <li>- Kontraindikationen für eine Kernspintomographie</li> <li>- Generalisierter, konvulsiver Status epilepticus innerhalb von vier Wochen vor der ersten, geplanten Dosis</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>(klinische und nicht nur EEG-basierte Diagnose)</p> <ul style="list-style-type: none"> <li>- Schwere Infektion (z.B. Pneumonie, Pyelonephritis oder Meningitis) vor der ersten, geplanten Dosis (eventuell späterer Einschluss)</li> <li>- Risiko ICV-bedingter Komplikationen wie z.B. Hydrozephalus oder Ventrikel-Shunt</li> <li>- Überempfindlichkeit gegenüber einem der Inhaltsstoffe des Prüfpräparats</li> <li>- Verabreichung eines anderen Prüfpräparates innerhalb von 30 Tagen vor der ersten Infusion der Studienmedikation oder geplante Gabe eines anderen Prüfpräparates im Verlauf der Studie</li> <li>- Andere medizinische Gründe oder Bedingungen, die aus Sicht des Prüfarztes gegen einen Einschluss des Patienten sprechen</li> <li>- Schwangerschaft</li> </ul> <p><b>Rekrutierung</b> Patienten, die den Eignungskriterien entsprachen, wurden konsekutiv durch die Zentren rekrutiert.</p> <p><b>Datenerhebung und Orte, an denen die Daten erhoben wurden</b> Die Studie fand in fünf Zentren statt: Universitätsklinikum Hamburg-Eppendorf, Hamburg, Deutschland, Great Ormond Street Hospital, London UK, Nationwide Children's Hospital, Columbus, USA, Evelina London Children's Hospital, London, UK und Bambino Gesù Children's Hospital, Rom, Italien.</p> <p>Der erste Patient wurde am 13.09.2013 in die Studie eingeschlossen und am 30.11.2015 endete die Studie.</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
4	<p>Intervention</p> <p>Details der vorgesehenen Interventionen für jede Studiensituation, wie und wann sie verabreicht wurden, und zwar einschließlich:</p> <ul style="list-style-type: none"> <li>• Inhalt: was wurde gegeben?</li> <li>• Verabreichungsmethode: wie wurde der Inhalt gegeben?</li> <li>• Verabreichungseinheit: wie wurden die Patienten während der Verabreichung eingeteilt?</li> <li>• Verabreicher: wer verabreichte die Intervention?</li> <li>• Umfeld: wo wurde die Intervention verabreicht?</li> <li>• Expositionsmenge und -dauer: wie viele Sitzungen oder Episoden oder Ereignisse sollten verabreicht werden? Wie lange sollten sie dauern?</li> <li>• Zeitspanne: wie lange sollte die Verabreichung der Intervention bei jeder Einheit dauern?</li> <li>• Aktivitäten zur Erhöhung der Compliance oder Adhärenz (z. B. Anreize)</li> </ul>	<p>Das Prüfpräparat ist Brineura® (Cerliponase alfa) Lösung zur intrazerebroventrikulären (ICV) Infusion. Pro Milliliter enthält die Lösung 30 mg Cerliponase alfa (entsprechend 150 mg pro 5 ml der Lösung bzw. pro Ampulle). Bei Cerliponase alfa handelt es sich um rekombinante, humane TPP1, die in Säugetier-CHO-Zellen hergestellt wird.</p> <p>Zur Verabreichung des Prüfpräparates wurden den Patienten ein ICV-Reservoir und eine ICV-Kanüle operativ implantiert. Cerliponase alfa wurde alle zwei Wochen jeweils morgens nachdem die Patienten mindestens zwei Stunden gefastet hatten ICV verabreicht (im Falle einer nächtlichen enteralen Ernährung wurde diese mindestens zwei Stunden vor der Infusion beendet). Während der Dosisescalation erhielten die Patienten 30, 100 und 300 mg Cerliponase alfa über jeweils mindestens vier Wochen. Die stabile Dosis betrug 300 mg alle zwei Wochen.</p> <p>Bei Dosen unter 300 mg wurde die Lösung auf 10 ml verdünnt. Die Flussrate betrug 2,5 ml/Stunde bei einer Infusionsdauer von etwa vier Stunden. Durch eine mit einem Alarm ausgestattete Pumpe wurde sichergestellt, dass das Prüfpräparat korrekt infundiert wurde.</p> <p>Die Patienten wurden vor jeder Infusion stationär aufgenommen. Während und nach den Infusionen wurden die Patienten engmaschig im Krankenhaus überwacht (mindestens 48 Stunden nach der ersten Dosis und jeder Dosiserhöhung [24 Stunden in der pädiatrischen Intensivstation] und bis zu 48 Stunden mindestens aber 24 Stunden während der stabilen Phase, ausgenommen die erste Dosis).</p> <p>Die Behandlungsdauer betrug mindestens 48 Wochen (länger bei Patienten, die an der Dosisescalation teilnahmen). Eine Follow-up-Visite zur Überprüfung der Sicherheit war sechs Monate nach der letzten Infusion vorgesehen.</p> <p>Das Prüfpräparat durfte nur durch qualifiziertes Krankenhauspersonal verabreicht werden. Das Datum, die Dauer, das Volumen und die Konzentration der Infusionen waren zu dokumentieren. Inkomplette oder fehlende Infusionen waren</p>



Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>in den CRFs unter Angabe der Gründe und anderer relevanter Informationen anzugeben. Da ausgewählte Mitarbeiter der Studienzentren das Prüfpräparat verabreichten, waren Compliance-Probleme nicht zu erwarten.</p>
5	<p>Zielsetzungen Spezifische Studienziele und Hypothesen</p>	<p><b>Primäres Studienziel</b></p> <ul style="list-style-type: none"> <li>- Sicherheit und Verträglichkeit von Cerliponase alfa bei Patienten mit CLN2</li> <li>- Wirksamkeit von Cerliponase alfa verglichen mit einer unbehandelten historischen Kontrollgruppe nach 12 Monaten (gemessen mittels einer CLN2-spezifischen Bewertungsskala, ML-Skala)</li> </ul> <p><b>Sekundäre Studienziele</b></p> <ul style="list-style-type: none"> <li>- Effekt der Behandlung auf die Gehirnatrophie verglichen mit einer unbehandelten, historischen Kontrollgruppe nach 12 Monaten</li> <li>- Charakterisierung der Pharmakokinetik nach Einmal- und Mehrfachgabe (CSF und Plasma)</li> <li>- Immunogenizität (CSF und Plasma)</li> </ul> <p><b>Explorative Studienziele</b></p> <ul style="list-style-type: none"> <li>- Effekt von Cerliponase alfa auf krankheitsspezifische CSF- und Plasma-Biomarker</li> <li>- Dokumentation altersgerechter Entwicklungsstufen</li> <li>- Krankheitsspezifische Lebensqualität</li> </ul> <p><b>Hypothesen:</b></p> <p>Primärer Endpunkt war die Änderung des 0-6-Punkt ML-Scores. Nach Daten zum natürlichen Verlauf schreitet die Erkrankung im Mittel innerhalb von 18 Monaten um 2 Punkte voran (SD 1,8). Mittels Einstichproben-t-Test wurde überprüft, ob die mittlere Krankheitsprogression (Abfall des ML-Scores) unter Behandlung mit Cerliponase alfa (<math>\mu</math>BMN-190) diesem Wert entspricht oder nicht:</p> <p>H0: <math>\mu</math>BMN-190 = 2.0  H1: <math>\mu</math>BMN-190 <math>\neq</math> 2.0 [51].</p> <p>Auf Empfehlung von Zulassungsbehörden wurde die oben dargestellte primäre Hypothese im Verlauf der Studie in eine Responder-Analyse (SAP) inklusive eines historischen Vergleichs (ISE SAP) umgewandelt. Die entsprechenden Hypothesen sind unter Item 11 dargestellt.</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
6	<p>Zielkriterien</p> <ul style="list-style-type: none"> <li>• Klar definierte primäre und sekundäre Wirksamkeitsmessgrößen</li> <li>• Methoden zur Datenerhebung und Methoden zur Verbesserung der Qualität von Messungen</li> <li>• Informationen über validierte Instrumente wie z. B. psychometrische und biometrische Eigenschaften von Selbsterhebungsinstrumenten</li> </ul>	<p><b>Primäre Zielparameter per ISE SAP Version 3 (25.04.2016) für die historischen Vergleiche</b> [58]</p> <p>Responder-Analyse: Änderung des 0-6-Punkt ML/HML-Scores (6 = altersadäquate Funktion bis 0 = weitgehend immobil und keine Sprache, Tabelle 4-28) nach 1:1 Paarbildung („matching“) mit einer historischen Kontrollgruppe. Ein Abfall um weniger als 2 Punkte/48 Wochen war als Ansprechen auf die Therapie definiert. Der Abfall des ML-Scores wurde von Baseline über einen Zeitraum von 48 Wochen projiziert.</p> <p>Kaplan Meier Analyse: Zeit bis zum ersten dauerhaften Abfall um 2 Punkte auf der ML/HML Skala.</p> <p>Unterschied der mittleren Rate der Abnahme auf der ML/HML-Skala.</p> <p><b>Sekundäre Zielkriterien</b></p> <p>Imputation der ML/HML Scores auf gemeinsame nominelle Zeitpunkte und deskriptive Statistik.</p> <p><b>Primärer Zielparameter nach Studienprotokoll 190-201 und SAP Version 3 (25.03.2016)</b></p> <p>Responder-Analyse: Anteil der Patienten, die auf die Therapie mit Cerliponase alfa angesprochen hatten, gemessen mittels 0-6-Punkt ML-Skala (ITT).</p> <p>Ein Ansprechen auf die Therapie war wie folgt definiert: Kein Abfall des ML-Wertes um zwei Punkte oder Abfall um zwei Punkte mit anschließendem Anstieg (kein stabiler Abfall) oder Wert von 0 auf der ML-Skala in Woche 48. Ein so genannter stabiler Abfall entsprach einem Abfall um zwei oder mehr Punkten, der bis zum letzten</p>

		<p>Untersuchungszeitpunkt nicht wieder um einen Punkt (oder mehr) zugenommen hatte. Die Ergebnisse wurden mit Daten zum natürlichen Verlauf der Erkrankung verglichen.</p> <p>Kaplan Meier Analyse: Zeit bis zum dauerhaften Abfall um 2 Punkte auf der ML Skala.</p> <p>Mittleren Rate der Abnahme auf der ML-Skala.</p> <p><b>Sekundäre Zielkriterien</b></p> <ul style="list-style-type: none"> <li>- MRI-Analyse (Hirnatrophie): <ul style="list-style-type: none"> <li>Gesamtvolumen des Gehirns (mm<sup>3</sup>)</li> <li>Volumen der CSF (mm<sup>3</sup> und % des Hirnvolumens)</li> <li>Volumen der kortikalen grauen Substanz (mm<sup>3</sup> und % des Hirnvolumens)</li> <li>Gesamtvolumen der weißen Substanz (mm<sup>3</sup> und % des Hirnvolumens)</li> <li>Scheinbarer Diffusionskoeffizient (mm<sup>2</sup>/Sekunde)</li> </ul> </li> <li>- Pharmakokinetik</li> <li>- Immunogenizität</li> </ul> <p><b>Explorative Zielkriterien</b></p> <ul style="list-style-type: none"> <li>- CSF- und Plasma-Biomarker (entfällt, da keine Biomarker identifiziert werden konnten)</li> <li>- Denver II Entwicklungs-Skala</li> <li>- Lebensqualität (PedsQL™ – Parent for Toddlers / Parent Family Impact, CLN2-spezifisches Instrument)</li> </ul> <p><b>Sicherheit und Verträglichkeit:</b></p> <p>UEs einschließlich SUEs, Überempfindlichkeitsreaktionen, Infusionsbedingte Reaktionen und UEs besonderen Interesses, Laboruntersuchungen (inklusive klinische Chemie, Hämatologie, Urinuntersuchung und CSF), Vitalzeichen, EKGs, EEGs und Immunogenität</p> <p>Studienspezifische SUEs wurden ab dem Zeitpunkt der Einverständniserklärung und alle anderen UEs/SUEs ab dem Zeitpunkt der ICV-Implantation dokumentiert. Zusätzlich sollten die UEs getrennt nach Zeitpunkt ihres Auftretens ausgewertet werden (nach der ICV-Implantation aber vor der ersten Infusion und nach der ersten Infusion).</p> <p>Die Daten wurden elektronisch unter Einsatz einer validierten internetbasierten Technologie dokumentiert.</p>
--	--	--

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Informationen zur Datenerhebung und Validität der Messinstrumente sind in Kapitel 4.4.2 und den jeweiligen Tabellen zur Operationalisierung der Endpunkte dargestellt (Tabelle 4-29 und Tabelle 4-39). Alle Prüfer nahmen an einer Schulung durch einen ausgewiesenen Experten teil und mussten sich vor der Studie qualifizieren. Während der Studie wurde regelmäßig überprüft, ob der ML-Score in allen Zentren korrekt und konsistent erhoben wurde. Den Prüfzentren wurde zusätzlich ein ausführliches Manual mit Angaben zur Durchführung und Bewertung der Tests zur Verfügung gestellt.</p>
7	<p>Fallzahl</p> <p>Wie die Stichprobengröße bestimmt wurde und, falls zutreffend, Erläuterung etwaiger Zwischenanalysen und Abbruchregeln</p>	<p>Nach Längsschnittdaten zum natürlichen Verlauf schreitet die CLN2 im Mittel um 2 Punkte (SD 1,8) innerhalb von 18 Monaten nach Auftreten erster Symptome voran (gemessen mittels HML-Skala). Unter der Annahme, dass die Erkrankung bei vergleichbarer Variabilität unter Behandlung um nur 0,5 Punkte voranschreitet, ergab sich eine Fallzahl von 18 auswertbaren Patienten (Power 90%, zweiseitiger Typ I-Fehler 0,05). Unter Berücksichtigung einer Abbruchrate von etwa 20% sollten 22 Patienten in die Studie eingeschlossen werden.</p> <p>Die Studie durfte nur aus Sicherheitsgründen aufgrund klinischer Kriterien vorzeitig abgebrochen werden; inferenzstatistische Abbruchregeln wurden dabei nicht eingesetzt. Während der Dosisescalation (Patienten Nr. 1-9) wurden die Sicherheitsdaten auf jeder Stufe analysiert und durch ein unabhängiges Datenkontrollkomitee bewertet. Im Studienprotokoll waren Bedingungen festgelegt (z.B. mit der Behandlung in Zusammenhang stehende CTCAE Grad 3 oder 4 Überempfindlichkeitsreaktionen), unter denen das Datenkontrollkomitee das Nutzen/Risiko-Verhältnis der Studie erneut zu bewerten hatte.</p>
8	<p>Zuweisungsmethode</p> <ul style="list-style-type: none"> <li>• Zuweisungseinheit (die Einheit, die der Studienbedingung zugewiesen wird, wie z. B. Person, Gruppe, Gemeinde)</li> <li>• Methode der Zuweisung von Einheiten zu Studienbedingungen, einschließlich Details</li> </ul>	<p>Da es sich um eine nicht kontrollierte Studie handelt, entfällt eine Zuteilung zu Behandlungsgruppen.</p> <p>Geeignete Patienten wurden konsekutiv nach Erscheinen rekrutiert.</p> <p>Die ersten 9 Patienten, die die Eignungskriterien der Studie erfüllten, wurden konsekutiv Kohorten aus jeweils drei Patienten zugeteilt und nahmen an einer</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
	<p>über etwaige Restriktion (z. B. Blockbildung, Stratifikation, Minimierung)</p> <ul style="list-style-type: none"> <li>• Einbeziehung von Aspekten, um potenzielle Verzerrung durch Nichtrandomisierung zu minimieren (z. B. Matching)</li> </ul>	<p>Dosiseskalation teil (ein Patient wurde zusätzlich eingeschlossen, da ein Patient die Studie vorzeitig abgebrochen hatte). Diese Patienten erhielten initial Dosen von 30, 100 bzw. 300 mg Cerliponase alfa ICV alle zwei Wochen über jeweils mindestens vier Wochen. Die nächst höhere Dosis wurde nur verabreicht, wenn bei der niedrigeren Dosis keine, nicht akzeptable Toxizität (entsprechend vordefinierter Kriterien) aufgetreten war. Über die Dosiseskalationen entschieden die Vertreter eines unabhängigen Datenkontrollkomitees auf Basis der Sicherheitsdaten. Nach Abschluss der Dosiseskalation wurde allen Patienten einschließlich der Patienten, die nicht an der Dosiseskalation teilgenommen hatten, eine stabile Dosis von 300 mg Cerliponase alfa (oder die höchste tolerierte Dosis) alle zwei Wochen über mindestens 48 Wochen verabreicht.</p> <p>Per SAP ISE (25.04.2016) wurde auf Empfehlung von Zulassungsbehörden ein historischer Vergleich des primären Endpunkts mit einer unbehandelten Kontrollgruppe als Primäranalyse festgelegt. Um das Risiko potenzieller Verzerrungen zu minimieren, wurden Kriterien zur Paarbildung (matching) der Studienteilnehmer und der historischen Kontrollen prospektiv definiert (primäre 1:1 Paarbildung sowie andere Algorithmen zur Überprüfung der Konsistenz; siehe Item 10).</p>
9	<p>Verblindung</p> <p>Unabhängig davon, ob sie Teilnehmer waren:</p> <ul style="list-style-type: none"> <li>• Wurden diejenigen, die die Interventionen durchführten, und diejenigen, die die Ergebnisse beurteilten, gegenüber der Zuweisung der Studienbedingung verblindet?</li> <li>• Wenn dies der Fall ist, Erklärung, wie die Verblindung erfolgte und wie sie bewertet wurde</li> </ul>	<p>Die Studie war nicht verblindet. Die Daten zur Sicherheit und Wirksamkeit wurden nicht verblindet erhoben.</p> <p>Die MRI-Daten wurden zentral durch unabhängige Radiologen ausgewertet, die keine Kenntnis des Patienten oder der jeweiligen Studiendauer hatten.</p>

<p><b>10</b></p>	<p>Analyseeinheit</p> <ul style="list-style-type: none"> <li>• Bezeichnung der kleinsten Einheit, die analysiert wird, um Wirkungen der Intervention zu beurteilen (z. B. Person, Gruppe, Gemeinde)</li> <li>• Wenn sich die Analyseeinheit von der Zuweisungseinheit unterscheidet, ist die verwendete Analysenmethode anzugeben, die dies berücksichtigt (z. B. Adjustierung der Standardfehlerschätzungen um den Designeffekt oder durch Verwendung einer Multilevel-Analyse)</li> </ul>	<p><b>Analysepopulationen per ISE SAP Version 3 – historische Verleiche - (25.04.2016) [58]</b></p> <p><u>Analyseeinheiten (historischer Vergleich):</u></p> <p>Die ITT-Population der Studie 190-201 wurde ausgewertet. Zur Überprüfung der Konsistenz wurden verschiedene, präspezifizierte Sensitivitätsanalysen eingesetzt.</p> <p>Auswertbare Patienten der Studie 190-901 (historische Kontrolle): Alle Patienten, die die Haupteinschlusskriterien der Studie 190-201 erfüllten, wurden berücksichtigt (vorausgesetzt sie entsprachen den jeweiligen Algorithmen zur Paarbildung; siehe unten): Alter <math>\geq 3</math> Jahre und mindestens ein ML-Score von <math>\geq 3</math> im Alter von <math>\geq 3</math> Jahren. In der Regel wurden nur Patienten der Studie 190-901 eingeschlossen, bei denen mindestens zwei ML-Werte im Bereich von 1-5 Punkten im Abstand von mindestens sechs Monaten vorlagen.</p> <p><u>Algorithmen zur Paarbildung</u> zwischen den Patienten der Studie 190-201 und den historischen Kontrollen:</p> <p>Es wurden verschiedene Algorithmen prospektiv eingesetzt.</p> <p>Die primäre Analyseeinheit umfasste diejenigen Patienten der ITT-Population, die einem Algorithmus nach 1:1 Paarbildung (matching) entsprachen (Methode #1). Der Algorithmus hatte zum Ziel, möglichst viele Patienten der Studie 190-201 den historischen Kontrollen zuzuteilen. Bei dem Algorithmus wurden lediglich Baseline-Kriterien berücksichtigt:</p> <ul style="list-style-type: none"> <li>- Exakt übereinstimmender ML/HML-Skalenwert</li> <li>- Möglichst hohe Übereinstimmung hinsichtlich des Alters, nicht mehr als 12 Monate Unterschied</li> </ul> <p>Bei bei Patienten der Studie 190-901 (historische Kontrolle) entsprach Baseline der Erhebung zum Zeitpunkt (Alter) der Paarbildung. Es wurde der jeweils längste Analysezeitraum bezogen auf Baseline berücksichtigt, der Analysezeitraum durfte allerdings nicht länger als derjenige des gepaarten (matched) Patienten der Studie 190-201 sein (mindestens sechs Monate bei Patienten der Studie 190-901).</p>
------------------	---	---

		<p><u>Weitere Algorithmen zur Paarbildung:</u></p> <p>Methode #1 zur 1:1 Paarbildung (matching), Simulationen: Die Analyse umfasste 1.000 Simulationen zur Paarbildung.</p> <p>Methode #2 zur 1:1 Paarbildung (matching): Der Algorithmus hatte ebenfalls eine maximale Paarbildung zum Ziel und beruhte ausschließlich auf Baseline-Kriterien. Der Algorithmus ähnelte der Methode #1, es wurde allerdings ein unterschiedliches Abstandsmaß verwendet (Summe aus dem Abstand zwischen Baseline-ML-Wert und Alter [<math>\leq 36</math> Monate] im Quadrat, möglichst niedriger Abstand):</p> $\sqrt{\left[\frac{901 \text{ Score} - 201 \text{ BL Score}}{1/6}\right]^2 + [901 \text{ Age} - 201 \text{ BL Age}]^2}$ <p>Zur Berechnung des Unterschieds der ML/HML-Scores (Baseline) wurde zunächst ein Faktor von 6 eingesetzt, da der Score basierend auf Daten zum natürlichen Verlauf im Mittel um 2 Punkte/48 Wochen abnimmt (entsprechend 1 Punkt/6 Monate). Analog zur Methode #1 entsprach das Alter der historischen Kontrollen zum Zeitpunkt der Paarbildung (matching) Baseline. Der Analysezeitraum entsprach ebenfalls Methode #1 (nicht länger als derjenige des Patienten der Studie 190-201, mindestens 6 Monate).</p> <p>„Many-to-one“ Paarbildung (matching) unter Berücksichtigung folgender Variablen:</p> <ul style="list-style-type: none"> <li>- ML/HML-Score zu Baseline,</li> <li>- ML/HML-Score und das Alter jeweils zu Baseline und</li> <li>- ML/HML-Score zu Baseline und Genotyp (c.622&gt;T und/oder c.509-1G&gt;C vorhanden oder nicht).</li> </ul> <p>Diese Analysen erlaubten - verglichen mit der 1:1 Paarbildung - den Einschluss einer größeren Zahl Patienten der Studie 190-901. Zusätzlich wurden die ITT-Population der Studie 190-201 und die auswertbaren Patienten der Studie 190-901 ohne Paarbildung (matching) analysiert.</p> <p><b>Analysepopulationen per SAP Version 3 – Studie 190-210 - (25.03.2016)</b></p> <p>Die ITT-Population umfasste alle Patienten mit Wirksamkeitsdaten, die (unabhängig von der Dosis) Cerliponase alfa erhalten hatten. Ein Patient, der die Studie nach nur einer Infusion abgebrochen hatte (Entzug der Einverständniserklärung), wurde aus allen</p>
--	--	--

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Wirksamkeits-Analysen einschließlich ITT ausgeschlossen [51].</p> <p>Die PP-Population umfasste in Einklang mit der ICH E9-Richtlinie alle ITT-Patienten, die den Vorgaben des Studienprotokolls entsprachen [51]. (Da die PP-Population der ITT-Population entsprach, entfiel die PP-Analyse.)</p> <p>In die Analyse zur Sicherheit wurden alle Patienten, denen ein ICV-Zugang implantiert worden war, eingeschlossen.</p>



11	<p>Statistische Methoden</p> <ul style="list-style-type: none"> <li>• Verwendete statistische Methoden zum Vergleich der Studiengruppen in Bezug auf die primäre(n) Wirksamkeitsvariable(n), einschließlich komplexer Methoden für korrelierte Daten</li> <li>• Verwendete statistische Methoden für zusätzliche Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen</li> <li>• Methoden zur Imputation fehlender Daten</li> <li>• Falls verwendet: Verwendete statistische Software oder Programme</li> </ul>	<p><i>Analyse des primären Endpunkts per ISE SAP Version 3 – historische Vergleiche - (25.04.2016; 1:1 matching)</i></p> <p><u>Responder-Analyse</u></p> <p>Auf Empfehlung der Zulassungsbehörden wurde die Responder-Analyse (ML/HML-Score, Abfall um &lt;2 Punkte/48 Wochen) geändert und durch einen historischen Vergleich mit Paarbildung (matching) ersetzt.</p> <p>Die folgende Hypothese wurde unter Einsatz des exakten Tests nach Fisher überprüft (behandelt vs. unbehandelt):</p> <p>H0: <math>p_{190-201/202} = p_{190-901}</math>  H1: <math>p_{190-201/202} \neq p_{190-901}</math></p> <p>(<math>p_{190-201/202}</math> = Patienten [in %], die auf Cerliponase alfa angesprochen hatten; <math>p_{190-901}</math> = Patienten der Studie 190-901, die den Responder-Kriterien entsprachen).</p> <p>Darüber hinaus wurden die motorische und Sprachdomäne des ML/HML-Scores wie oben beschrieben getrennt analysiert (Abfall um &lt;1 Punkt pro 48 Wochen).</p> <p>In der Primäranalyse entsprach Baseline der letzten Erhebung vor der ersten 300 mg-Infusion („300 mg-Baseline“; Studie 190-201). Bei Patienten der Studie 190-901 war Baseline im SAP kontextabhängig definiert:</p> <ul style="list-style-type: none"> <li>- 1:1 Paarbildung (matching): Erhebung zum Zeitpunkt der Paarbildung (ausgedrückt als Alter in Monaten)</li> <li>- Many-to-one Paarbildung (matching): Identifikation aneinandergrenzender Visiten mit dokumentierten ML-Werten, die zu den Baseline-Werten der Patienten der Studie 190-201 passten (matched). Dies konnten eine einzelne oder mehrere Visiten sein. Der mittlere Wert zwischen der ersten und letzten Visite entsprach Studientag 1.</li> <li>- Analysen ohne Paarbildung (Matching): Identifikation des ersten HML-Wertes unter 6 Punkten (ausgedrückt als Alter in Monaten). Falls mehrere aneinandergrenzende Werte vorlagen, wurde wie oben beschrieben vorgegangen.</li> </ul> <p><i>Sensitivitäts(Subgruppen)-Analysen:</i></p> <ul style="list-style-type: none"> <li>- Die verschiedenen Algorithmen zur Paarbildung sind unter Item 10 beschrieben.</li> <li>- Ausschluss von Patienten mit einem ML/HML-Skalenwert von 6 Punkten zu Baseline, der während der Studie stabil blieb. Diese Patienten hatten</li> </ul>
----	--	---

		<p>wahrscheinlich die Phase der aktiven Krankheitsprogression („decline phase“, d.h. ML/HML-Score &lt;6) noch nicht erreicht.</p> <ul style="list-style-type: none"> <li>- Ausschluss von Patienten mit einem ML/HML-Skalenwert &lt;3 zu Baseline. Diese Analyse war aus Gründen der Konsistenz mit den Einschlusskriterien der Studie 190-201 sinnvoll. Bei einigen Patienten (190-201) nahm der ML-Score von <math>\geq 3</math> zur Screening-Visite auf &lt;3 zur Baseline-Visite ab (das Einschlusskriterium bezog sich auf die Screening-Visite).</li> <li>- Ausschluss von Patienten mit einem ML/HML-Skalenwert von 6, der während der Studie stabil blieb oder Baseline-Wert &lt;3.</li> </ul> <p><u>Abnahme des ML-Scores über 48 Wochen</u> Der Endpunkt (Zeit bis zum ersten stabilen Abfall um 2 Punkte oder Wert von 0 wurde mittels Kaplan-Meier und eines Cox-Modells untersucht. In dem Cox-Modell wurde nach ML/HML-Score und Alter zu Baseline adjustiert.</p> <p>Unter Einsatz eines Zweistichpropen-t-Tests wurde der mittlere Abfall bei behandelten und unbehandelten Patienten verglichen (zweiseitig auf einem Signifikanzniveau von <math>\alpha=0,05</math>). Die folgenden Hypothesen wurden überprüft:  <math>H_0: \mu_{190-201/202} = \mu_{190-901}</math>  <math>H_1: \mu_{190-201/202} \neq \mu_{190-901}</math>  (<math>\mu_{190-201/202}</math> = Mittelwert der Abnahme unter Cerliponase alfa; <math>\mu_{190-901}</math> = Mittelwert der Abnahme in der Kontrollgruppe).</p> <p>Zur Primäranalyse wurde die 1:1 Paarbildung (matching) nach Methode #1 eingesetzt.</p> <p><i>Unterstützende und Sensitivitätsanalysen</i> Andere Paarbildungen sowie eine Analyse ohne Paarbildung wurden unterstützend eingesetzt. Die Robustheit der Primäranalyse wurde in oben genannten Sensitivitätsanalysen (Ausschluss stabiler ML/HML-Score von 6 mit oder ohne Baseline-Score &lt;3) überprüft.</p> <p>Die motorische und Sprachkomponente wurden ebenfalls getrennt analysiert (Methode #1). Die Abnahme des MLV-Scores (ML plus Sehvermögen, 0-9 Punkte) wurde mittels „many-to-one“ Paarbildung sowie ohne Paarbildung untersucht. Zusätzlich wurde ein MMRM-Modell zur</p>
--	--	--

		<p>Analyse des Abfalls des ML/HML- und MLV-Scores eingesetzt (ITT ohne Paarbildung; Random-Slope- und Intercept-Modell).</p> <p>Die Verteilung der ML/HML-Slopes bei behandelten und unbehandelten Patienten wurde unter Einsatz einer kumulativen Verteilungsfunktion untersucht.</p> <p><u>Fehlende Werte</u></p> <p>Weder die Analysen der Krankheitsprogression (Slopes) noch die Ereigniszeitanalysen erforderten eine Imputation fehlender Werte.</p> <p>Die HML-Daten in der Studie 190-901 nicht wie in 190-201 regelmäßig erhoben. Verschiedene Imputationsmethoden wurden eingesetzt, um zu erreichen, dass der Datensatz für jeden Patienten beider Studien zu bestimmten Zeitpunkten (Analysetage 169, 253, 337 und 505) Werte enthielt:</p> <ul style="list-style-type: none"> <li>- Vorzugsweise wurde eine lineare Interpolation eingesetzt. Lag lediglich ein Einzelwert innerhalb von 28 Tagen bezogen auf den Analysetag vor, wurde dieser berücksichtigt.</li> <li>- Eingabe desjenigen Werts, der dem Zeitpunkt am nächsten lag („windowing“).</li> </ul> <p>Bei Patienten (Studie 190-901), bei denen sich der HML-Score während aufeinanderfolgender Visiten nicht änderte, wurde angenommen, dass der Score zwischen den Visiten stabil blieb. Für die Analysen mit 1:1 Paarbildung (matching) wurde zwischen diesen Visiten monatlich der entsprechende HLM-Wert eingegeben.</p> <p>Der ursprüngliche Datensatz enthielt abgeleitete (imputierte) Visiten, die identifiziert und entfernt wurden.</p> <p><b><i>(Ursprüngliche) Primäranalyse (SAP Version 3, 25.03.2016)</i></b></p> <p><u>Responder-Analyse</u></p> <p>In der primären Analyse wurde der Anteil der Patienten in Prozent, die in Woche 48 auf die Behandlung angesprochen hatten (0-6-Punkt ML-Skala), untersucht. Die Wirksamkeit von Cerliponase alfa wurde mittels Einstichprobentest (exakter Binomialtest) wie folgt überprüft:</p> <p>H0: <math>p_{BMN-190} \leq 0,5</math>  H1: <math>p_{BMN-190} &gt; 0,5</math></p>
--	--	---

		<p>(pBMN-190 = Anteil Patienten in Prozent, die auf Cerliponase alfa angesprochen hatten)</p> <p>Als Sensitivitätsanalyse wurde der Anteil Patienten (in %) analysiert, bei denen sich der ML-Score in Woche 48 nicht stabil um einen Punkt verschlechtert hatte. Die Hypothesengrenze betrug 0,25 (exakter Binomialtest).</p> <p>Die Hypothesengrenzen (0,50 bzw. 0,25) wurden wie folgt bestimmt: Bei unbehandelten Patienten nimmt der Score im Mittel um 2 Punkte/48 Wochen mit einer SD von 1 ab. Vorausgesetzt die Werte sind normalverteilt, nimmt der Wert erwartungsgemäß bei 50% der unbehandelten Patienten um weniger als 2 Punkte und bei 16% um weniger als 1 Punkt ab. Daher sind die Null-Hypothesen (Ansprechrates <math>\leq 0,50</math> bzw. <math>\leq 0,25</math>) angemessen.</p> <p><u>Abnahme des ML-Scores über 48 Wochen</u></p> <p>Zur Analyse wurde der Einstichproben-t-Test eingesetzt und folgende Hypothese wurde überprüft:</p> <p><math>\mu_{\text{BMN-190}} \geq 2,0</math>  <math>\mu_{\text{BMN-190}} &lt; 2,0</math></p> <p>Die Hypothesengrenze basiert auf Daten zum natürlichen Verlauf bei unbehandelten CLN2-Patienten (siehe oben). Algorithmus zur Berechnung der Rate des Abfalls:</p> $\text{Slope} = \frac{(\text{Finaler ML-Score}) - (\text{ML-Score zu Beginn})}{(\text{Finaler Zeitpunkt}) - (\text{Zeitpunkt zu Beginn})}$ <p>Rate des Abfalls = <math>(-1) \times (48 \times 7) \times \text{Slope}</math></p> <p><u>Sensitivitäts(Subgruppen)-Analysen</u></p> <p>Ausschluss von Patienten mit einem Baseline ML-Score von 6,</p> <ul style="list-style-type: none"> <li>- der während der Studie nicht abnimmt</li> <li>- der während der Studie nicht abnimmt oder Baseline ML-Score von <math>&lt; 3</math></li> </ul> <p><u>Sekundäre Endpunkte</u></p> <p>Die sekundären Endpunkte wurden deskriptiv analysiert.</p> <p><u>Subgruppen- und andere Analysen (fakultativ)</u></p> <ul style="list-style-type: none"> <li>- Genotypen insbesondere von c.622C&gt;T und c.509-1G&gt;C Mutationen</li> <li>- Demographische Charakteristika, Studienzentrum oder geographische Region</li> <li>- Unterschiedliche Behandlungsintervalle wie gesamtes Dosisintervall, nach Einleitung der 300 mg-Dosis, während</li> </ul>
--	--	--

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>der stabilen Dosierung oder nach Verabreichung einer bestimmten Menge des Prüfpräparates (z.B. 90 Tage oder nach 1.000 mg Cerliponase alfa)</p> <p>Diese Analysen waren explorativ.</p> <p><u>Imputation fehlender Werte</u></p> <p>Fehlende Werte der primären Responder-Analyse wurden durch Einsatz der Kaplan-Meier-Methode ersetzt. Andere Imputationen waren nicht vorgesehen.</p>
-	<b>Ergebnisse</b>	
12	<p>Patientenfluss</p> <ul style="list-style-type: none"> <li>• Durchlauf der Studienteilnehmer durch jede Studienphase: Rekrutierung, Zuweisung, Allokation und Interventionsexposition, Nachbeobachtung, Analyse (ein Diagramm wird dringend empfohlen)</li> <li>• Rekrutierung: die Anzahl der Teilnehmer, die auf ihre Eignung gescreent werden, sich als geeignet oder nicht geeignet erweisen, nicht in die Studie aufgenommen und in die Studie aufgenommen werden</li> <li>• Zuweisung: die Anzahl der Teilnehmer, die einer Studienbedingung zugewiesen wurden</li> <li>• Allokation und Interventionsexposition: die Anzahl der Teilnehmer, die jeder Studienbedingung zugewiesen werden, und die Anzahl der Teilnehmer, die jede Intervention erhielten</li> <li>• Nachbeobachtung: die Anzahl der Teilnehmer, die das Follow-up abschlossen oder das Follow-up nicht abschlossen (lost to follow-up), aufgeschlüsselt nach Studienbedingungen</li> <li>• Analyse: die Anzahl der Teilnehmer, die in die Hauptanalyse einbezogen oder ausgeschlossen wurden, aufgeschlüsselt nach Studienbedingungen</li> <li>• Beschreibung von Abweichungen vom Prüfplan der Studie unter Angabe von Gründen</li> </ul>	<p><b><i>Patientenfluss (Studie 190-201):</i></b></p> <p>Eingeschlossene Patienten: 24 (Dosisescalation: 10, direkter Einschluss in die stabile Phase: 14)</p> <p>Behandelte Patienten: 24 (Analyse zur Sicherheit und Sensitivitätsanalysen)</p> <p>Patienten, die die Studie abgeschlossen haben: 23 (ITT)</p> <p>Alle geeigneten Patienten wurden bis zum Erreichen der festgelegten Fallzahl in die Studie eingeschlossen. Ein Patient der Kohorte 3 zog nach nur einer Infusion sein Einverständnis zurück und wurde durch einen weiteren Patienten ersetzt. Bei den verbleibenden 23 Patienten wurde nur eine Infusion nicht verabreicht.</p> <p>Patienten, die die Studie 190-201 abgeschlossen hatten, konnten an einer offenen Folgestudie teilnehmen. Alle 23 Patienten nehmen weiterhin an dieser teil. Der Patientenfluss ist in Abbildung 20 dargestellt.</p> <p><b><i>Beschreibung von Abweichungen vom Prüfplan unter Angabe der Gründe</i></b></p> <p>Der erste Patient wurde am 13.09.2014 unter Amendment 2 des Studienprotokolls (26.07.2013) eingeschlossen.</p> <p>Wichtige Änderungen werden nachfolgend dargestellt (43):</p> <p><u>Amendment 1 (10.05.2013):</u></p> <p>Nachbeobachtung nach sechs Monaten statt vier Wochen aufgrund präklinischer Daten</p> <p>Aus Sicherheitsgründen mindestens 72 Stunden zwischen der ersten Dosis bei</p>

		<p>dem ersten und zweiten Patienten (Kohorte 1)</p> <p>Klarstellung der Notwendigkeit einer Vorbehandlung mit Anti-Histaminen und der Möglichkeit zusätzlicher Antipyretika</p> <p>Auf Empfehlung einer europäischen Ethik-Kommission Ausschluss von Patienten im Alter von 16 Jahren oder älter</p> <p>Aus Sicherheitsgründen Ausschluss von weiblichen Patienten im gebärfähigen Alter</p> <p><u>Amendment 2 (26.07.2013):</u></p> <p>Amendment 2 war bereits vor Einschluss der ersten Patienten in Kraft.</p> <p>Darstellung der Rationale, die dem Volumen und der Infusionsrate des Prüfpräparates zugrunde liegt</p> <p>Informationen zur Sedierung während der Infusion</p> <p>Informationen zur Überwachung epileptischer Anfälle während der Infusion und der Möglichkeit einer Unterbrechung der Infusion</p> <p>Notwendigkeit einer regelmäßigen Überprüfung der Funktion des ICV-Zugangs</p> <p>Diese Änderungen entsprachen Vorgaben von EU-Behörden und/oder lokaler Ethik-Kommissionen.</p> <p><u>Amendment 3 (07.07.2014):</u></p> <p>Änderung der Zeitpunkte der MRIs, um das mit einer Allgemeinanästhesie und Sedation verbundene Risiko zu vermindern</p> <p>Videoaufnahme aller Erhebungen der adaptierten CLN2-Skala zur Überprüfung und Ergänzung der klinischen Daten (ursprünglich alle 24 Wochen)</p> <p>Spezifizierung der Dauer der stationären Überwachung der Patienten nach Implantation des ICV-Zugangs, der ersten Dosis und bei Dosiserhöhung (48 Stunden, davon 24 Stunden in der pädiatrischen Intensivstation); die regelmäßige Überwachung der Studie machte diese Änderung erforderlich.</p> <p>Änderung der Fallzahl von 22 auf etwa 22 Patienten, um Patienten, die die Studie abbrechen, ersetzen zu können</p> <p>Der Prüfarzt sollte bewerten, ob ein UE mit dem ICV-Zugang in Zusammenhang stand oder nicht, um eine bessere Einschätzung dieser UEs zu ermöglichen</p>
--	--	---

		<p>Klarstellung, dass alle Screening-Untersuchungen innerhalb von <math>\leq 3</math> Tagen vor der Operation (ICV) abgeschlossen sein müssen</p> <p>Zusätzliche Bestimmung der gesamten und Cerliponase alfa-spezifischen IgE zu Baseline, um die Ergebnisse bei Auftreten einer Überempfindlichkeitsreaktion besser bewerten zu können</p> <p><u>Amendment 4 (08.08.2014):</u> Spezifizierung der zu berichtenden UEs, die mit den Infusionsmaterialien in Zusammenhang stehen (alle Komponenten, die mit Cerliponase alfa in Kontakt kommen)</p> <p>Informationsmaterial zu möglichen Komplikationen des ICV-Zugangs für Patienten und deren Betreuer</p> <p>Mit diesen Änderungen wurde den Empfehlungen von Behörden entsprochen.</p> <p><u>Amendment 5 (24.09.2014):</u> Weitere Spezifizierung von UEs, die mit den Infusionsmaterialien in Zusammenhang stehen (alle Komponenten)</p> <p>Engmaschige stationäre Überwachung der Patienten über bis zu 48 Stunden während der stabilen Dosierung, allerdings durften klinisch stabile Patienten nach 24 Stunden entlassen werden, vorausgesetzt es bestanden keine Sicherheitsbedenken.</p> <p>Mit diesen Änderungen wurde Empfehlungen von Behörden entsprochen.</p> <p><u>SAP Version 2.0:</u> Nach SAP Version 1.0 sollten Patienten mit einem ML-Score von 3-6 und einem Score von mindestens 1 auf jeder der Domänen unmittelbar vor der ersten 300 mg-Dosis in die Primäranalyse eingeschlossen. Dies war per SAP Version 2.0 nicht mehr erforderlich (die ursprüngliche Population wurde in einer Sensitivitätsanalyse ausgewertet). Die Primäranalyse sollte dem ITT-Prinzip und der in der ISE analysierten Population besser entsprechen[51].</p> <p><u>SAP Version 3.0 (25.03.2016):</u> Nach Diskussionen mit den Zulassungsbehörden wurde eine Responder-Analyse als primärer Endpunkt definiert; die ursprüngliche Analyse blieb allerdings als wichtiger Endpunkt erhalten [51].</p> <p>Änderung der Zielsetzung: Immunogenität explorativer Endpunkt im Protokoll und sekundärer Endpunkt im SAP; QoL sekundärer Endpunkt im Protokoll und explorativer Endpunkt im SAP (bei Diskrepanzen galten die Angaben im SAP)</p>
--	--	---

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p><u>SAP ISE Version 3 (25.04.2016):</u> In Übereinstimmung mit den Empfehlungen der Zulassungsbehörden wurde die primäre Responder-Analyse um ein 1:1 Matching der ITT-Population der Studien 190-201/202 und der auswertbaren Patienten der Studie 190-901 ergänzt (sowie zusätzliche Paarbildungs-Algorithmen; siehe oben). Die Slope-Analyse blieb weiterhin als bedeutender Endpunkt erhalten. Die Hypothesentestung wurde von einem Einstichprobentest in einen Zweistichprobentest geändert. Es wurde klargestellt, dass sich die Wirksamkeitsanalyse auf das 300 mg-Dosisintervall der Studien 190-201/202 bezieht (Baseline: letzter Wert vor der ersten 300 mg-Dosis).</p> <p><u>Studienbericht (finaler Datenschnitt) [26]:</u></p> <p>Auf die Analyse der kumulativen Verteilung der Änderung (Slope) der ML-Scores wurde verzichtet, da diese gut bekannt ist und viele Nullwerte vorlagen.</p> <p>Der Einfluss einer 12-monatigen Therapie mit Cerliponase alfa auf die Gehirnatrophie (sekundäres Studienziel) wurde nicht wie im Protokoll vorgegeben mit historischen Daten zum natürlichen Verlauf verglichen. Es liegen keine MRI-Daten der Studie 190-901 vor.</p>
13	Aufnahme / Rekrutierung Daten, die den Rekrutierungs- und den Nachbeobachtungszeitraum eingrenzen	Der erste Patient wurde am 13.09.2013 in die Studie aufgenommen und die Studie wurde am 30.11.2015 protokollgemäß beendet. Alle Patienten, die die Studie abgeschlossen hatten, nehmen an der Folgestudie 190-202 teil.
<b>a: nach TREND 2004.</b>		



Tabelle 4-76 (Anhang): Studiendesign und -methodik für Studie 190-202 (fortlaufende Studie) [45]

Item <sup>a</sup>	Charakteristikum	Studieninformation
-	<b>Studienziel</b>	
2	<p>Hintergrund / Rationale</p> <ul style="list-style-type: none"> <li>Wissenschaftlicher Hintergrund und Erläuterung der Rationale</li> <li>Theorien bei der Konzeption von Verhaltensinterventionen (nicht zutreffend)</li> </ul>	<p>Zur Behandlung der CLN2 stehen bislang lediglich unterstützende Maßnahmen zur Verfügung und es besteht bei dieser sehr schwerwiegenden Erkrankung ein sehr hoher, bislang ungedeckter therapeutischer Bedarf. Es wird erwartet, dass Cerliponase alfa als ERT das bei CLN2-Patienten fehlende Enzym TPP1 ersetzt und dadurch die krankheitsrelevante Speicherung lysosomalen Materials vermindert.</p> <p>Der hohe therapeutische Bedarf sowie die sehr vielversprechenden Ergebnisse tierexperimenteller Studien rechtfertigen die klinische Entwicklung von Cerliponase alfa. Bei Bassets mit einer TPP1-Nullmutation verlängerte Cerliponase alfa (ICV) signifikant die Zeit bis zum Auftreten von Symptomen, erhielt die motorischen und kognitiven Fähigkeiten, verminderte die Hirnatrophie und verlängerte das Leben der Hunde [74]. Die Effekte waren besonders deutlich, wenn die Therapie möglichst früh nach der Geburt eingeleitet wurde.</p> <p>Die Ergebnisse der klinischen Vorläuferstudie 190-201 zeigten eine insgesamt gute Verträglichkeit von Cerliponase alfa und bei den meisten Patienten (15 von 23) blieb die Erkrankung stabil oder besserte sich (2 Patienten) unter Behandlung mit Cerliponase alfa.</p>
-	<b>Methoden</b>	
3	<p>Studienteilnehmer</p> <ul style="list-style-type: none"> <li>Eignungskriterien für Studienteilnehmer, einschließlich der Kriterien auf verschiedenen Ebenen des Rekrutierungsplans / Plan zur Stichprobennahme (sampling plan) (z. B. Städte, Kliniken, Patienten)</li> <li>Rekrutierungsmethode (z. B. Überweisung, Selbstauswahl) einschließlich der Methode zur Stichprobennahme (sampling method), falls ein systematischer Plan zur Stichprobennahme verwendet wurde</li> <li>Vorgehensweise bei der Rekrutierung</li> <li>Setting der Datenerhebung und Orte, an denen die Daten erhoben wurden</li> </ul>	<p><b>Einschlusskriterien:</b></p> <ul style="list-style-type: none"> <li>Abschluss der Studie 190-201 (48 Wochen).</li> <li>Schriftliche Einverständniserklärung</li> <li>Sexuelle Abstinenz oder Einsatz einer hocheffektiven Verhütungsmethode bei sexuell aktiven Männern und Frauen (bis zu sechs Monate nach der Studie oder nach einem Abbruch)</li> <li>Negativer Schwangerschaftstest bei weiblichen Patienten im gebärfähigen Alter</li> </ul> <p><b>Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>Abfall des ML-Scores um <math>\geq 3</math> Punkte während der Studie 190-201 (Baseline bis zur letzten Visite) und der Patient</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>profitiert nach Ansicht des Prüfarztes nicht von der Therapie</p> <ul style="list-style-type: none"> <li>- ML-Score von Null</li> <li>- Schwangerschaft oder Stillzeit oder der/die Patient(in) oder deren/dessen Partner(in) plant eine Schwangerschaft während der Studie</li> <li>- Anwendung anderer Prüfparate oder investigativer Medizinprodukte innerhalb von 30 Tagen vor oder während der Studie</li> <li>- Begleiterkrankungen oder andere Bedingungen, die nach Ansicht des Prüfarztes gegen eine Teilnahme sprechen oder ein Risiko darstellen</li> <li>- Möglicherweise schlechte Compliance oder hohes Risiko eines Studienabbruchs aus Sicht des Prüfarztes</li> </ul>
4	<p>Intervention</p> <p>Details der vorgesehenen Interventionen für jede Studiensituation, wie und wann sie verabreicht wurden, und zwar einschließlich:</p> <ul style="list-style-type: none"> <li>• Inhalt: was wurde gegeben?</li> <li>• Verabreichungsmethode: wie wurde der Inhalt gegeben?</li> <li>• Verabreichungseinheit: wie wurden die Patienten während der Verabreichung eingeteilt?</li> <li>• Verabreicher: wer verabreichte die Intervention?</li> <li>• Umfeld: wo wurde die Intervention verabreicht?</li> <li>• Expositionsmenge und –dauer: wie viele Sitzungen oder Episoden oder Ereignisse sollten verabreicht werden? Wie lange sollten sie dauern?</li> <li>• Zeitspanne: wie lange sollte die Verabreichung der Intervention bei jeder Einheit dauern?</li> <li>• Aktivitäten zur Erhöhung der Compliance oder Adhärenz (z. B. Anreize)</li> </ul>	<p>Den Patienten wird alle zwei Wochen vorzugsweise am Morgen 300 mg Cerliponase alfa ICV verabreicht (2,5 ml/Stunde über etwa 4 Stunden unter Einsatz einer Spritzenpumpe). Das Prüfpräparat soll appliziert werden nachdem die Patienten mindestens zwei Stunden gefastet haben.</p> <p>Die Patienten werden vor jeder Cerliponase alfa-Infusion stationär aufgenommen. Cerliponase alfa wird nur von gut ausgebildeten Krankenhausangestellten verabreicht und eine Notfallausrüstung steht für den Fall einer allergischen Reaktion bereit. Die Patienten werden stationär über bis zu 48 Stunden engmaschig überwacht. Falls keine Sicherheitsbedenken bestehen, kann der Patient nach 24 Stunden entlassen werden.</p> <p>Die geplante Behandlungsdauer beträgt bis zu 239 Wochen.</p> <p>Da ausgewählte Mitarbeiter der Studienzentren das Prüfpräparat verabreichen, sind Compliance-Probleme nicht zu erwarten.</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
5	Zielsetzungen Spezifische Studienziele und Hypothesen	<p><b>Primäre Studienziele:</b></p> <ul style="list-style-type: none"> <li>- Langzeitsicherheit von Cerliponase alfa 300 mg alle zwei Wochen bei Patienten mit CLN2</li> <li>- Änderung der ML-Skala unter Behandlung mit Cerliponase alfa</li> </ul> <p><b>Sekundäre Studienziele</b></p> <ul style="list-style-type: none"> <li>- Änderung quantitativer Parameter im MRI</li> <li>- Änderung des krankheitsspezifischen CLN2-Gesamtscores</li> <li>- QoL unter Langzeitbehandlung mit Cerliponase alfa</li> </ul> <p><b>Explorative Studienziele</b></p> <ul style="list-style-type: none"> <li>- Erreichen altersgemäßer Entwicklungsschritte („Meilensteine“) unter Langzeitbehandlung mit Cerliponase alfa</li> <li>- Einfluss der Behandlung auf krankheitsspezifische Biomarker in der CSF oder im Plasma</li> </ul>
6	Zielkriterien <ul style="list-style-type: none"> <li>• Klar definierte primäre und sekundäre Wirksamkeitsmessgrößen</li> <li>• Methoden zur Datenerhebung und Methoden zur Verbesserung der Qualität von Messungen</li> <li>• Informationen über validierte Instrumente wie z. B. psychometrische und biometrische Eigenschaften von Selbsterhebungsinstrumenten</li> </ul>	<p><b>Sicherheit</b></p> <ul style="list-style-type: none"> <li>- UEs und Begleitmedikation</li> <li>- Routine-Laborparameter</li> <li>- Routine-CSF-Tests</li> <li>- Vitalzeichen</li> <li>- Körperliche Untersuchungen</li> <li>- EKG</li> <li>- Immunogenität</li> </ul> <p><b>Wirksamkeit</b></p> <ul style="list-style-type: none"> <li>- Krankheitsspezifische CLN2-Bewertungsskala unter Videoüberwachung</li> <li>- Volumen der kortikalen, grauen Substanz</li> <li>- QoL-Parameter</li> <li>- Entwicklung der Patienten (explorativ)</li> <li>- Biomarker (explorativ)</li> <li>- Neurologische Untersuchungen</li> </ul> <p><b>Immunogenität</b></p> <ul style="list-style-type: none"> <li>- Anti-Cerliponase alfa-Antikörper (gesamt und neutralisierend) in der CSF und Anti-Cerliponase alfa-Antikörper (gesamt) sowie Anti-Cerliponase alfa-IgE im Serum</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
7	Fallzahl Wie die Stichprobengröße bestimmt wurde und, falls zutreffend, Erläuterung etwaiger Zwischenanalysen und Abbruchregeln	Eine Fallzahlschätzung war nicht erforderlich, da es sich um eine Folgestudie handelt, an der alle Patienten, die die Studie 190-201 abgeschlossen hatten, teilnehmen konnten. Es wurde angenommen, dass 23 Patienten an der Studie teilnehmen würden. Die Studiendaten werden in regelmäßigen Abständen von BioMarin zusammenfassend analysiert. Ein vorzeitiger Abbruch der Studie ist nur aus Gründen der Sicherheit auf Basis einer klinischen Bewertung möglich.
8	Zuweisungsmethode <ul style="list-style-type: none"> <li>• Zuweisungseinheit (die Einheit, die der Studienbedingung zugewiesen wird, wie z. B. Person, Gruppe, Gemeinde)</li> <li>• Methode der Zuweisung von Einheiten zu Studienbedingungen, einschließlich Details über etwaige Restriktion (z. B. Blockbildung, Stratifikation, Minimierung)</li> <li>• Einbeziehung von Aspekten, um potenzielle Verzerrung durch Nichtrandomisierung zu minimieren (z. B. Matching)</li> </ul>	Patienten, die 48 Wochen in der Studie 190-201 abgeschlossen hatten, konnten an der Studie teilnehmen. Es ist geplant, die Ergebnisse zum Krankheitsverlauf (CLN2-Score) mit Verlaufsdaten bei unbehandelten Patienten aus mindestens zwei Registern (WCMC und Hamburg) zu vergleichen.
9	Verblindung Unabhängig davon, ob sie Teilnehmer waren: <ul style="list-style-type: none"> <li>• Wurden diejenigen, die die Interventionen durchführten, und diejenigen, die die Ergebnisse beurteilten, gegenüber der Zuweisung der Studienbedingung verblindet?</li> <li>• Wenn dies der Fall ist, Erklärung, wie die Verblindung erfolgte und wie sie bewertet wurde</li> </ul>	Die Studie ist nicht verblindet. Die Daten zur Sicherheit und Wirksamkeit werden nicht verblindet erhoben. Die MRI-Scans werden zentral durch unabhängige Radiologen ausgewertet, denen der Patient und die Behandlungsdauer unbekannt sind.
10	Analyseeinheit <ul style="list-style-type: none"> <li>• Bezeichnung der kleinsten Einheit, die analysiert wird, um Wirkungen der Intervention zu beurteilen (z. B. Person, Gruppe, Gemeinde)</li> <li>• Wenn sich die Analyseeinheit von der Zuweisungseinheit unterscheidet, ist die verwendete Analysenmethode anzugeben, die dies berücksichtigt (z. B. Adjustierung der Standardfehlerschätzungen um den Designeffekt oder durch Verwendung einer Multilevel-Analyse)</li> </ul>	In die Analyse werden die Daten der Studie 190-201 sowie der Studie 190-202 eingeschlossen. Die Analyseeinheiten entsprechen dem SAP der Studie 190-201 und sind in Tabelle 4-75 beschrieben.
11	Statistische Methoden <ul style="list-style-type: none"> <li>• Verwendete statistische Methoden zum Vergleich der Studiengruppen in Bezug auf die primäre(n) Wirksamkeitsvariable(n),</li> </ul>	Der SAP der Studie 190-201 gilt gleichermaßen für die Studie 190-202. Die Analysen sind in Tabelle 4-75 beschrieben. Die Daten werden unter Einsatz von SAS <sup>®</sup> analysiert.

Item <sup>a</sup>	Charakteristikum	Studieninformation
	<p>einschließlich komplexer Methoden für korrelierte Daten</p> <ul style="list-style-type: none"> <li>• Verwendete statistische Methoden für zusätzliche Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen</li> <li>• Methoden zur Imputation fehlender Daten</li> <li>• Falls verwendet: Verwendete statistische Software oder Programme</li> </ul>	
-	<b>Ergebnisse</b>	
12	<p>Patientenfluss</p> <ul style="list-style-type: none"> <li>• Durchlauf der Studienteilnehmer durch jede Studienphase: Rekrutierung, Zuweisung, Allokation und Interventionsexposition, Nachbeobachtung, Analyse (ein Diagramm wird dringend empfohlen)</li> <li>• Rekrutierung: die Anzahl der Teilnehmer, die auf ihre Eignung gescreent werden, sich als geeignet oder nicht geeignet erweisen, nicht in die Studie aufgenommen und in die Studie aufgenommen werden</li> <li>• Zuweisung: die Anzahl der Teilnehmer, die einer Studienbedingung zugewiesen wurden</li> <li>• Allokation und Interventionsexposition: die Anzahl der Teilnehmer, die jeder Studienbedingung zugewiesen werden, und die Anzahl der Teilnehmer, die jede Intervention erhielten</li> <li>• Nachbeobachtung: die Anzahl der Teilnehmer, die das Follow-up abschlossen oder das Follow-up nicht abschlossen (lost to follow-up), aufgeschlüsselt nach Studienbedingungen</li> <li>• Analyse: die Anzahl der Teilnehmer, die in die Hauptanalyse einbezogen oder ausgeschlossen wurden, aufgeschlüsselt nach Studienbedingungen</li> <li>• Beschreibung von Abweichungen vom Prüfplan der Studie unter Angabe von Gründen</li> </ul>	<p><b>Zwischenanalyse: Datenschnitt 15.10.2015</b> [27]</p> <p>Insgesamt 19/23 Patienten, die die Studie 190-201 abgeschlossen hatten, wurden bis zum Datenschnitt eingeschlossen.</p> <p>Die verbleibenden fünf Patienten schlossen die Studie 190-201 nach dem Datenschnitt ab und wurden am 1.12.2015 in der Studie aufgenommen.</p> <p>Die Patienten waren zum Zeitpunkt der Analyse im Mittel über 16,0 zusätzliche Wochen (0-30,7 Wochen) behandelt worden (nur Studie 190-202).</p> <p>Alle Patienten wurden in die Analysen eingeschlossen (n=23 ITT).</p> <p><b>Zwischenanalyse (auf Anforderung der FDA): Datenschnitt Juni 2016</b> [34]</p> <p>Alle 23 Patienten nehmen weiterhin an der Studie teil und kein Patient hat die Studie 190-202 vorzeitig abgebrochen.</p> <p>Zum Zeitpunkt der Zwischenanalyse betrug die mittlere, kumulative Behandlungsdauer 96,4 (SD 29,77) Wochen (Studie 190-201 plus 190-202).</p> <p>Alle Patienten wurden in die Analysen eingeschlossen (n=23 ITT).</p> <p><u>Amendment 3 des Studienprotokolls (26.02.2016):</u></p> <ul style="list-style-type: none"> <li>- Änderung der Anwendung von Verhütungsmethoden während der Studie und der Nachbeobachtung auf Anforderung einer Behörde</li> <li>- Keine CSF-Probenahme während der Nachbeobachtung; stattdessen Überprüfung der Immunogenizität im Serum</li> </ul>
13	<p>Aufnahme / Rekrutierung</p> <p>Daten, die den Rekrutierungs- und den Nachbeobachtungszeitraum eingrenzen</p>	<p>Der erste Patient wurde am 02.02.2015 in die Studie aufgenommen. Die Studie ist noch nicht abgeschlossen und die letzte Visite findet erwartungsgemäß im Dezember 2020 statt.</p>
<b>a: nach TREND 2004.</b>		



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

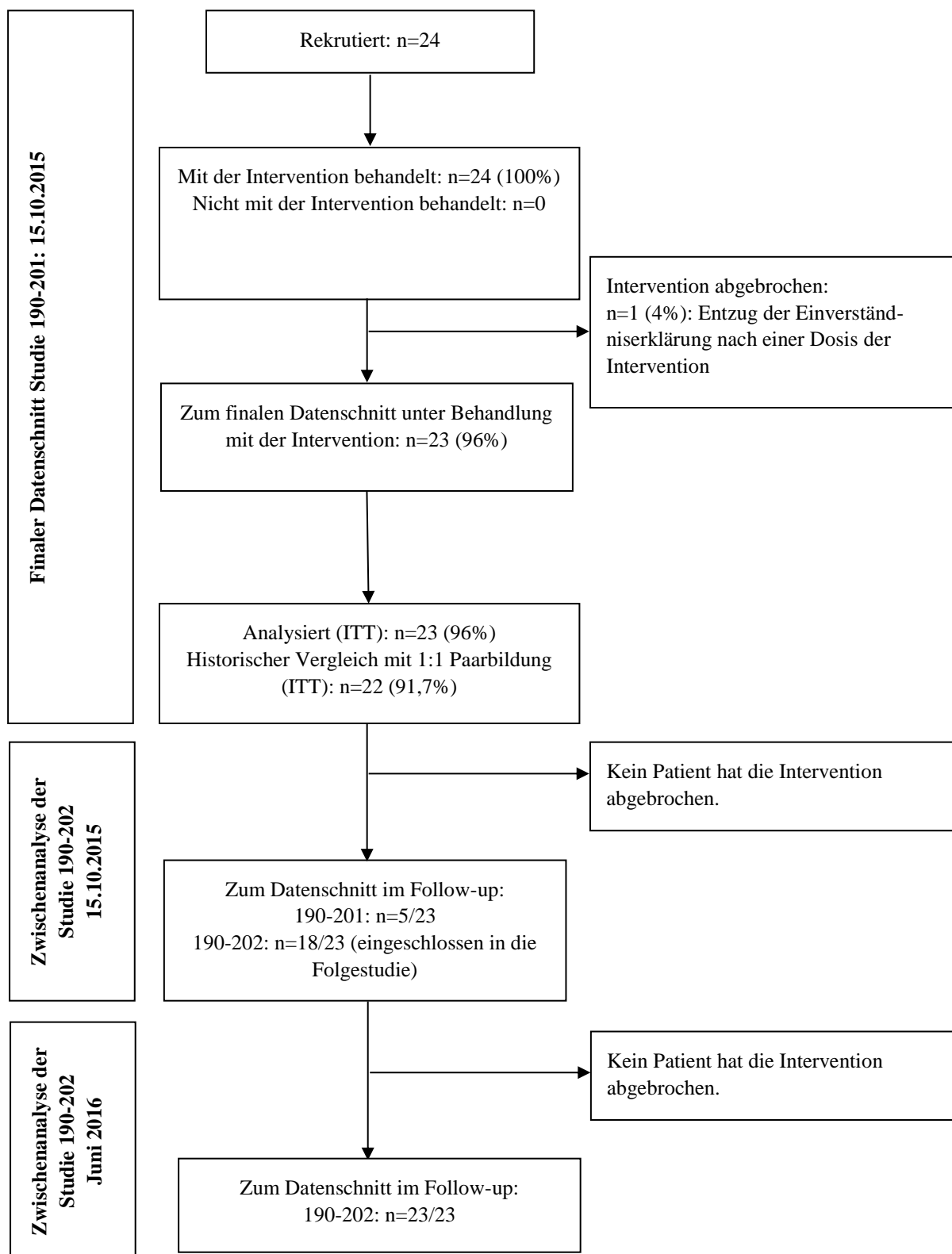


Abbildung 20: Patientenfluss der Studie 190-201 und der Folgestudie 190-202

Tabelle 4-77 (Anhang): Studiendesign und -methodik der Studie 190-901 (190-901 Supplemental Report zum natürlichen Verlauf der CLN2) und des historischen Vergleichs mit den Studien 190-201/202 [56]

Item <sup>a</sup>	Charakteristikum	Studieninformation
-	<b>Titel und Abstract</b>	
	<b>Einleitung</b>	
2	Hintergrund/Rationale	<p><b>Hintergrund</b></p> <p>Bei der CLN2 handelt es sich um eine sehr seltene Erkrankung und es steht kein zugelassenes Arzneimittel zur Verfügung. Zur Überprüfung der Wirksamkeit von Cerliponase alfa ist eine aktiv kontrollierte Studie daher nicht möglich und eine placebokontrollierte Studie schließt sich aufgrund der besonderen Schwere der Erkrankung und des erforderlichen ICV-Zugangs aus ethischen Gründen aus. Eine Parallelgruppe unbehandelter Patienten war aufgrund der Seltenheit der Erkrankung nur schwer bis nicht realisierbar. Darüber hinaus bestehen ethische Bedenken, Kinder, deren Erkrankung häufig dramatisch und irreversibel fortschreitet, von der Behandlung auszuschließen. Um die Wirksamkeit von Cerliponase alfa bei CLN2 nachzuweisen, wurde daher ein historischer Vergleich (matched) herangezogen.</p> <p>Das Hauptziel der Studie 190-901 ist die Beschreibung des natürlichen Krankheitsverlaufs bei einem größeren Patientenkollektiv, um die Wirksamkeit von Cerliponase alfa vs. eine historische Kontrolle überprüfen zu können [56].</p>
3	Zielsetzungen	<ul style="list-style-type: none"> <li>- Analyse des Krankheitsverlaufs bei unbehandelten Patienten mit CLN2, um als historische Kontrolle in Cerliponase alfa-Studien zu dienen</li> <li>- Analyse der Krankheitsprogression und deren Variabilität nach Auftreten von Symptomen unter Einsatz einer krankheitsspezifischen Skala bei einem größeren Kollektiv</li> <li>- Beschreibung der Krankheitscharakteristika mittels deskriptiver Methoden</li> </ul> <p>Die Analysen waren explorativer Natur und es wurden dem Design der Studie entsprechend keine spezifischen Hypothesen überprüft.</p>
-	<b>Methoden</b>	
4	Studiendesign	<p>Retrospektive Analyse der Längsschnittdaten einer klinischen Datenbank (DEM-CHILD). Retrospektiv bedeutet in diesem Zusammenhang, dass die in der Datenbank vorhandenen Angaben nachträglich analysiert wurden. Zum Zeitpunkt der Einrichtung der Datenbank war die Analyse nicht geplant. Zudem enthält die Datenbank teilweise</p>



Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>retrospektive Angaben (Daten, die nicht prospektiv durch klinische Beobachtung erhoben wurden).</p> <p>Der ursprüngliche DEM-CHILD Datensatz enthielt imputierte Visiten, die wahrscheinlich auf retrospektiven Angaben seitens der Eltern/Behandler vor der Diagnosestellung basierten. Nach Diagnosestellung wurde der HML-Score prospektiv durch klinische Beobachtung erhoben (Original-Analyse der Studie 190-901, [39]). Wie im ISE SAP [58] festgelegt wurden die imputierten Visiten soweit möglich vor der Analyse aus dem Datensatz entfernt.</p>
5	Rahmen (Setting)	<p>Die Datenbank DEM-CHILD ist eine fortlaufende, multizentrische, multinationale, klinische Datenbank mit Zentrum in Hamburg. In der Datenbank werden klinische, Labor- und bildgebende Daten sowie Informationen zur Entwicklung von Patienten mit NCLs (u. a. CLN2) erfasst. DEM-CHILD stellt eine Erweiterung des ursprünglichen Datensatzes aus Hamburg dar (1. Report Studie 190-901 [39]), der Daten zum natürlichen Verlauf der Erkrankung enthält. Die Daten wurden in ausgewiesenen Zentren in zwei Krankenhäusern erhoben. Zum Zeitpunkt der Analyse (Februar 2015) lagen Daten von CLN2-Patienten aus zwei Zentren vor (Hamburg, n=63, Verona, n=11). Bei insgesamt 29 Patienten aus Hamburg handelte es sich um Patienten des Original-Datensatzes der Studie 190-901 [39].</p> <p>Es liegen lediglich Daten zum Geburtsjahr der Patienten vor (n=41 auswertbare Patienten), die Hinweise zum Zeitpunkt der Datensammlung liefern könnten. Insgesamt sechs Patienten (15%) wurden zwischen 1960 und 1989 geboren, 19 (46%) zwischen 1990 und 1999 und 16 (39%) zwischen 2000 und 2009.</p> <p>Aufgrund der Art der Datensammlung unterschieden sich die Beobachtungszeiträume zwischen den Patienten.</p>
6	Studienteilnehmer	<p><b>Einschlusskriterien</b></p> <p>In die Analyse wurden Patienten mit CLN2 eingeschlossen. Abgesehen von drei Patienten (7% der 41 auswertbaren Patienten) war die Diagnose bei allen Patienten genotypisch gesichert.</p> <p>Die in der Datenbank identifizierten Patienten mit CLN2 (n=74) wurden mittels verschiedener Filter selektiert, um ein der Studie 190-201 möglichst</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>ähnliches Patientenkollektiv zu erhalten. Folgende Filter wurden eingesetzt:</p> <ul style="list-style-type: none"> <li>- Ausschluss von Zwillingen</li> <li>- Patienten mit mindestens zwei HML-Scores <ul style="list-style-type: none"> <li>Zwei HML-Scores im Alter von <math>\geq 36</math> Monaten</li> <li>Ein HML-Score <math>\geq 3</math> Punkten</li> <li>Zwei HML-Scores zwischen 1 und 5 Punkten einschließlich</li> <li>Mindestens ein Wert <math>\geq 6</math> Monate nach der ersten HML-Bewertung</li> </ul> </li> </ul> <p>Insgesamt 41 Patienten entsprachen diesen Kriterien, von denen 33 aus Hamburg (Deutschland) und 8 aus Verona (Italien) stammten.</p> <p>Der Beobachtungszeitraum richtete sich nach den in der Datenbank verfügbaren Daten (mindestens 6 Monate hinsichtlich des ML-Scores).</p>
6.1	Methoden zur Selektion der Studienpopulation (RECORD Statement)	<p>Zur 1:1 Paarbildung von Patienten der Studie 190-201/201 und der historischen Kontrolle wurde der folgende Algorithmus zur Primäranalyse eingesetzt:</p> <ul style="list-style-type: none"> <li>- Exakt übereinstimmender ML/HML-Score zu Baseline und</li> <li>- Möglichst hohe Übereinstimmung hinsichtlich des Alters zu Baseline, nicht länger als 12 Monate</li> </ul> <p>In unterstützenden Analysen wurden andere Algorithmen zu Paarbildung eingesetzt. Es liegen zudem Analysen aller auswertbaren historischen Kontrollen vor (siehe Tabelle 4-75).</p>
6.2	Validierungsstudien der Codes oder Algorithm (RECORD Statement)	Eine Validierung der Algorithmen erscheint bei der gegenständlichen Fragestellung nicht erforderlich.
6.3	Verknüpfung von Datenbanken (RECORD Statement)	Nicht zutreffend
7	Variablen	Es wurden alle in der Datenbank bei CLN2-Patienten verfügbaren Daten, die für die

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Beantwortung der gegenständlichen Fragestellung relevant sind, extrahiert.</p> <p><u>Primärer Endpunkt:</u> Krankheitsprogression gemessen mittels HLM-Skala</p> <p><u>Andere Variablen und Endpunkte:</u></p> <ul style="list-style-type: none"> <li>- Demographische Parameter wie Geburtsdatum und Geschlecht soweit vorhanden</li> <li>- Art der ersten klinischen Symptome</li> <li>- Alter bei Auftreten der ersten klinischen Symptome gesamt und bei Patienten, bei denen u.a. Krampfanfälle aufgetreten waren</li> <li>- Alter bei Diagnosestellung</li> <li>- Symptome, die zur Diagnose führten</li> <li>- HML-Score bei (oder vor) Diagnosestellung</li> <li>- Genotypisierung und TPP-Assay (soweit Daten vorhanden)</li> </ul> <p>MRI-Daten wurden bei CLN2-Patienten nicht systematisch erfasst.</p>
8	Datenquellen/Messmethoden	<p>Die HML-Skala wurde erstmals 2002 durch eine Arbeitsgruppe in Hamburg beschrieben [5], die maßgeblich an DEM-CHILD beteiligt ist und ein wichtiges Zentrum der Studien 190-201/202 ist.</p> <p>Die Original-HML-Skala wurde in der Studie 190-901 und eine geringfügig modifizierte Version (ML-Skala) in den Studien 190-201/202 eingesetzt. Die Unterschiede zwischen beiden Skalen sind in Tabelle 4-29 dargestellt. Die Modifikationen basierten auf Empfehlungen von Experten und dienten dazu, eine konsistente Bewertung sowohl zwischen den Zentren als auch zeitlich zu gewährleisten. Aufgrund der nur minimalen Änderungen können beide Skalen als gleichwertig eingestuft werden.</p>
9	Bias	<p>Ein Selektionsbias erscheint unwahrscheinlich, da alle Patienten, bei denen in den Studienzentren während des Beobachtungszeitraums eine CLN2 diagnostiziert wurde, in die Analyse einbezogen wurden. Die zur Paarbildung mit Patienten der Studie 190-201 eingesetzten Algorithmen hatten zum Ziel, eine historische Kontrollgruppe zu erreichen, die den Haupteinschlusskriterien der Studie 190-201 soweit wie möglich entsprach. Die zur Paarbildung gewählten Algorithmen waren daher wissenschaftlich begründet. Die Konsistenz der Ergebnisse wurde in verschiedenen Sensitivitätsanalysen u.a. ohne Paarbildung überprüft (siehe Tabelle 4-75).</p> <p>Zur Imputation fehlender Werte wurden konservative Methoden eingesetzt. Die Konsistenz der Ergebnisse wurde mittels MMRM überprüft</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>(siehe Tabelle 4-75 zum historischen Vergleich der Studie 190-201).</p> <p>Aufgrund des Designs sind die Studie und deren Ergebnisse potenziell hoch verzerrt, allerdings stellt sie die derzeit beste, verfügbare Evidenz mit Längsschnittdaten zum Krankheitsverlauf bei CLN2 dar. Die zur Messung der Krankheitsprogression in den Studien eingesetzten Instrumente sind vergleichbar (HML-Skala, 190-201/202 ML-Skala). Die Ergebnisse der Studie 190-901 werden durch unabhängige Studien zum Krankheitsverlauf bestätigt (siehe z. B. [6], Abbildung 8).</p>
10	Studiengröße	<p>Die Studiengröße ergibt sich aus der Zahl der in den Zentren verfügbaren Patienten, unter Berücksichtigung der Seltenheit der Erkrankung. Das Studiendesign schließt eine Fallzahlschätzung aus.</p> <p>Es wurden zunächst alle in dem Datensatz identifizierten Patienten berücksichtigt. Anschließend fand eine Selektion entsprechend der Eignungskriterien statt, um eine den Studien 190-201/202 möglichst vergleichbare Patientenpopulation zu erzielen.</p>
11	Quantitative Variablen	<p>In der Analyse der Krankheitsprogression wurden Werte zwischen 5 und 1 auf der HML-Skala berücksichtigt. Ein Abfall des HML-Wertes von 6 auf 5 Punkte und 1 auf 0 Punkte wurde nicht berücksichtigt, da der Abfall in diesen Bereichen nicht linear ist. Zudem sollte die Phase des aktiven Fortschreitens der Erkrankung („decline phase“) analog zur Studie 190-201 erfasst werden. Patienten mit einem Score von 6 auf der HML-Skala sind in ihrem Geh- und Sprachvermögen nicht oder nur wenig eingeschränkt, während Patienten mit einem Score von Null weitgehend immobil sind und nicht mehr oder nur unverständlich sprechen können.</p> <p>Die Analyse des zur Beantwortung der gegenständlichen Fragestellung relevanten historischen Vergleichs inklusive der primären Responder-Analyse ist in Tabelle 4-75 beschrieben.</p>
12	Statistische Methoden	<p>Die folgenden Angaben beziehen sich auf die Analyse des DEM-CHILD-Datensatzes. Die statistische Analyse des historischen Vergleichs ist in Tabelle 4-75 beschrieben.</p> <p>Zur Kalkulation der klinischen Krankheitsprogression (Abfall des HML-Skalenwerts) über 48 Wochen (1 Monat = 30 Tage) wurden zwei Algorithmen eingesetzt:</p> <ul style="list-style-type: none"> <li>- Erster/Letzter Wert: Die Linie zwischen zwei HML-Skalenwerten vs. die Zeit wurde analysiert. Der erste Wert entsprach dem ersten HML-Skalenwert unter 6. Falls mehrere Werte unter 6 vorlagen, wurde der mittlere dieser</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Werte berücksichtigt. Der letzte Wert entsprach dem letzten Score &gt;0.</p> <p>- Regression: Die Linie zwischen dem ersten und letzten HML-Skalenwert wurde wie oben beschrieben analysiert. Alle HML-Scores dazwischen wurde mittels linearer Regression ausgewertet.</p> <p><u>Sensitivitätsanalyse: MMRM</u></p> <p>Die Daten wurden nach Alter (alle sechs Monate ab einem Alter von 36 Monaten) analysiert bis erstmals ein Wert von Null auftrat. Zur MMRM-Analyse wurde ein Model mit nicht strukturierter Varianz und linearem Trend eingesetzt. Der Einfluss des Alters zum Diagnosezeitpunkt wurde untersucht. Zusätzlich wurde ein Model mit autoregressiver Varianzstruktur zur Analyse aller 6-Monatsdaten eingesetzt.</p> <p>Da die HML-Scores nicht exakt alle sechs Monate erhoben wurden, wurde zunächst ein Fenster von <math>\pm 1</math> Monat für alle Werte nach dem Ausgangswert eingesetzt. Falls mehrere Werte innerhalb dieses Fensters vorhanden waren, wurde der Wert, der dem jeweiligen 6-Monatswert am nächsten lag (entweder vor oder nach dem Wert), eingesetzt. Für Fenster ohne Werte wurde die LOCF (last observation carried forward)-Imputation eingesetzt. Fehlende Ausgangswerte (Fenster <math>\pm 1</math> Monat) wurden in ähnlicher Weise gehandhabt, allerdings wurde der letzte gemessene Wert zurück versetzt (last observation carried backwards, LOCB). Die Daten wurden bis zu sechs Monate nach dem letzten Wert und bis zu sechs Monate vor dem ersten Wert unter Einsatz der LOCF- bzw. LOCB-Methode extrapoliert.</p> <p>In ähnlichen Modellen wurde das Alter zum Zeitpunkt der Diagnose (6, 12, 18 etc. Monate seit Diagnosestellung) untersucht.</p> <p>Zusätzlich wurde der Zeitraum untersucht, über den der HML-Score innerhalb von zwei benachbarten Punkten stabil blieb (z.B. 5/4, 4/3 etc.). Der Zeitraum war wie folgt definiert: Zeit zwischen dem ersten höheren Score des HML-Paares und dem ersten um zwei (ober mehr) Punkte niedrigeren Wert.</p>
12.1	Zugriff auf die Daten (RECORD Statement)	Keine Informationen vorliegend.
12.2	Datenbereinigung (RECORD Statement)	Keine Informationen vorliegend.
12.3	Verknüpfung von Daten (RECORD Statement)	Nicht zutreffend
	<b>Ergebnisse</b>	

Item <sup>a</sup>	Charakteristikum	Studieninformation
13	Teilnehmer	<p><b>Patientenfluss:</b></p> <p>Von den 74 in der Datenbank identifizierten Patienten mit CLN2 wurden 33 aus folgenden Gründen ausgeschlossen:</p> <p>n=1 identischer Zwilling</p> <p>n=15 nicht mindestens ein HML-Score</p> <p>n=7 nicht mindestens zwei HML-Scores</p> <p>n=2 kein HML-Score <math>\geq 3</math></p> <p>n=1 keine zwei HML-Scores zwischen 1 und 5</p> <p>n=7 nicht mindestens ein HML-Score <math>\geq 6</math> Monate nach der ersten Erhebung</p> <p>Insgesamt 41 Patienten erfüllten die Eignungskriterien, von denen 26/29 Patienten aus Hamburg bereits in dem ersten Bericht der Studie 190-901[39] analysiert worden waren. Bei 15 Patienten (7 Hamburg und 8 Verona) handelte es sich um neu aufgenommene Patienten.</p>
<p><b>a: nach STROBE 2007, 2008 unter Berücksichtigung des RECORD-Statements 2015.</b></p>		

**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 4-78 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie 190-201

**Studie: 190-201**

**Tabelle: Liste der für die Bewertung herangezogenen Quellen**

Genauere Benennung der Quelle	Kürzel
Finaler Studienbericht [26]	A
Studienprotokoll (Amendements A5) [37]	B
SAP Version 3 [51]	C
ISE SAP [58]	D
ISE [57]	E
190-901 Supplemental Report (historische Kontrolle) [56]	F

**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 einarmige, nicht kontrollierte Studie (B).

---

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:

---

Es handelt sich um eine einarmige, nicht kontrollierte Studie (B). Die Ergebnisse des primären Endpunkts wurden mit einer historischen Kontrollgruppe nach Paarbildung hinsichtlich prognostisch relevanter Faktoren verglichen (D,E).

---



2.

**für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“)** ja     unklar     neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

**für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** ja     unklar     neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Es handelt sich um eine einarmige, nicht kontrollierte Studie (B). Die Ergebnisse des primären Endpunkts wurden mit einer historischen Kontrollgruppe verglichen (D,E). Die Kriterien zur Paarbildung der Patienten der Studie 190-201 und der historischen Kontrollen entsprachen den Haupteinschlusskriterien der Studie 190-201, um hinsichtlich prognostisch relevanter Faktoren vergleichbare Gruppen zu erreichen; in einer Sensitivitätsanalyse wurde der Genotyp als zusätzliches Kriterium berücksichtigt (D,E).

---

3. **Verblindung von Patienten und behandelnden Personen****Patient:** ja     unklar     neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

Es handelt sich um eine offene Studie und die Patienten waren nicht verblindet (A).

---

**behandelnde bzw. weiterbehandelnde Personen:** ja     unklar     neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

Es handelt sich um eine offene Studie und die behandelnden bzw. weiterbehandelnden Personen waren nicht verblindet (A).

---

4. **Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte** ja     unklar     neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Es ergeben sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung der relevanten Endpunkte. Die Endpunkte wurden wie im finalen SAP bzw. ISE SAP festgelegt ausgewertet (C,D). Die Analyse des primären Endpunkts wurde im Verlauf der Studie geändert (B). Mit diesen Änderungen wurde Empfehlungen der Zulassungsbehörden

entsprochen, so dass es sich nicht um eine ergebnisgesteuerte Berichterstattung seitens BioMarin handelt.

#### 5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können

ja       nein

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

Es existieren keine Hinweise auf sonstige (endpunktübergreifende) Aspekte, die das Verzerrungspotenzial beeinflussen könnten. Geplante summarische Zusammenfassungen der Sicherheitsdaten zu Beginn der Studie dienten dem DMC als Entscheidungsgrundlage hinsichtlich der Dosisescalation (A,B). Diese waren aus Sicherheitsgründen unerlässlich, da es sich bei Cerliponase alfa um ein ERT handelt und die Studie 190-201 die erste Studie in Patienten war. Im September 2014 und Januar 2015 fanden nicht geplante Zwischenanalysen der Sicherheit und Wirksamkeit statt. Den Analysen lag die Intention zugrunde, einen vorzeitigen Antrag auf Zulassung zu stellen. Der Patientenfluss der Studie 190-201 ist nachvollziehbar und transparent. Es kann nicht vollständig ausgeschlossen werden, dass sich die Begleitbehandlungen zwischen der Studie 190-201 und der historischen Kontrollgruppe unterschieden. Zur Begleitbehandlung der historischen Kontrollgruppe liegen keine Angaben vor (F). Da es sich bei den üblicherweise eingesetzten Begleitbehandlungen um rein symptomatische/palliative Therapien handelt, die die motorischen Funktionen oder die Sprachfähigkeit nicht oder nur minimal beeinflussen, erscheint eine relevante Verzerrung des primären Endpunkts durch eine möglicherweise unterschiedliche Begleitbehandlung unwahrscheinlich. Zudem kann angenommen werden, dass die Begleitbehandlung im Rahmen einer prospektiven Interventionsstudie optimiert wird, sodass sich hinsichtlich der Wirksamkeit und Sicherheit mögliche Unterschiede der Begleitbehandlung tendenziell eher zugunsten von Cerliponase alfa verglichen mit der historischen Kontrollgruppe ausgewirkt hätten.

#### Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

niedrig       hoch

Begründung für die Einstufung:

Es handelt sich um eine nicht randomisierte Studie, sodass die Einstufung des Verzerrungspotenzials auf Studienebene entfällt.

#### B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:

##### Endpunkt: Änderung des ML-Scores

##### 1. Verblindung der Endpunkterheber

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Es handelt sich um eine offene Studie und der Endpunkt wurde nicht verblindet erhoben (A,B).

**2. Adäquate Umsetzung des ITT-Prinzips**

ja       unklar       nein

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

---

Das ITT-Prinzip wurde adäquat umgesetzt. Der Ausschluss eines Patienten aus der ITT-Population, der die Studie nach nur einer Dosis Cerliponase alfa abgebrochen hatte, war wissenschaftlich begründet (A).

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja       unklar       nein

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

---

Der Endpunkte wurden wie im finalen SAP bzw. ISE SAP festgelegt ausgewertet (C,D). Die Analyse des primären Endpunkts wurde im Verlauf der Studie geändert (B). Mit diesen Änderungen wurde Empfehlungen der Zulassungsbehörden entsprochen, so dass es sich nicht um eine ergebnisgesteuerte Berichterstattung seitens BioMarin handelt.

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja       nein

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

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig       hoch

Begründung für die Einstufung:

---

Es handelt sich um eine nicht randomisierte Studie, sodass die Einstufung des Verzerrungspotentials auf Endpunktebene entfällt.

---

**Endpunkt: PedsQL™****1. Verblindung der Endpunkterheber**

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Es handelt sich um eine offene Studie und der Endpunkt wurde nicht verblindet erhoben (A,B).

---

**2. Adäquate Umsetzung des ITT-Prinzips** ja     unklar     neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Das ITT-Prinzip wurde adäquat umgesetzt. Der Ausschluss eines Patienten aus der ITT-Population, der die Studie nach nur einer Dosis Cerliponase alfa abgebrochen hatte, war wissenschaftlich begründet (A).

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja     unklar     neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Der Endpunkte wurden wie im finalen SAP festgelegt deskriptiv ausgewertet (C).

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja     neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig     hoch

Begründung für die Einstufung:

---

Es handelt sich um eine nicht randomisierte Studie, sodass die Einstufung des Verzerrungspotentials auf Endpunktebene entfällt.

---

**Endpunkt: CLN2-spezifisches QoL-Instrument****1. Verblindung der Endpunkterheber** ja     unklar     neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

Es handelt sich um eine offene Studie und der Endpunkt wurde nicht verblindet erhoben (A,B).

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja     unklar     nein

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

---

Das ITT-Prinzip wurde adäquat umgesetzt. Der Ausschluss eines Patienten aus der ITT-Population, der die Studie nach nur einer Dosis Cerliponase alfa abgebrochen hatte, war wissenschaftlich begründet (A).

---

### 3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja     unklar     nein

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

---

Der Endpunkte wurden wie im finalen SAP festgelegt deskriptiv ausgewertet (C).

---

### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja     nein

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

### Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig     hoch

Begründung für die Einstufung:

---

Es handelt sich um eine nicht randomisierte Studie, sodass die Einstufung des Verzerrungspotentials auf Endpunktebene entfällt.

---

### Endpunkt: Sicherheit und Verträglichkeit

#### 1. Verblindung der Endpunkterheber

ja     unklar     nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Es handelt sich um eine offene Studie und der Endpunkt wurde nicht verblindet erhoben (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 „Sicherheits“-Analyse wurden alle Patienten, denen ein ICV-Zugang implantiert worden war, eingeschlossen (A).

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja       unklar       nein

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

---

Der Endpunkte wurden wie im finalen SAP festgelegt deskriptiv ausgewertet (C).

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja       nein

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

---

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig       hoch

Begründung für die Einstufung:

---

Es handelt sich um eine nicht randomisierte Studie, sodass die Einstufung des Verzerrungspotentials auf Endpunktebene entfällt.

Auf eine Darstellung des Verzerrungspotentials der Studie 190-202 auf Studien- sowie Endpunktebene wurde verzichtet, da die Studie nicht abgeschlossen ist und deren Zwischenergebnisse lediglich zur unterstützenden Information dargestellt wurden. Bei der Studie 190-202 handelt es sich um eine Folgestudie der Studie 190-201, deren Design weitestgehend mit demjenigen der Studie 190-201 übereinstimmt.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen mit Ausfüllhinweisen dient nur als Ausfüllhilfe für den Blankobogen. Er soll nicht als Vorlage verwendet werden.

### Bewertungsbogen zur Beschreibung von Verzerrungsaspekten (Ausfüllhilfe)

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunkübergreifend; B: endpunktspezifisch).

#### A Verzerrungsaspekte auf Studienebene:

##### Einstufung als randomisierte Studie

**ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien

**nein:** Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anzeichen vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

---

---

1.

##### für randomisierte Studien:

##### Adäquate Erzeugung der Randomisierungssequenz

**ja:** Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

**unklar:** Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

**nein:** Die Erzeugung der Zuteilungssequenz war nicht adäquat.

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

---

---

##### für nicht randomisierte Studien:

##### Zeitliche Parallelität der Gruppen

**ja:** Die Gruppen wurden zeitlich parallel verfolgt.

**unklar:** Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

**nein:** Die Gruppen wurden nicht zeitlich parallel verfolgt.

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

---

2.

**für randomisierte Studien:****Verdeckung der Gruppenzuteilung („allocation concealment“)** **ja:** Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

 **unklar:** Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau. **nein:** Die Gruppenzuteilung erfolgte nicht verdeckt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**für nicht randomisierte Studien:****Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** **ja:** Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anzeichen dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

 **unklar:** Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau. **nein:** Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Verblindung von Patienten und behandelnden Personen****Patient:** **ja:** Die Patienten waren verblindet. **unklar:** Es finden sich keine diesbezüglichen Angaben. **nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.



Angaben zum Kriterium; obligate Begründung für die Einstufung:

---



---

**behandelnde bzw. weiterbehandelnde Personen:**

**ja:** Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, die primär behandelnde Person (z. B. Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

**unklar:** Es finden sich keine diesbezüglichen Angaben.

**nein:** Aus den Angaben geht hervor, dass die behandelnden Personen nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---



---

**4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte**

*Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).*

*Beispiele zu a und b:*

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

*Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter*

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

*Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:*

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

*Zulässige Gründe sind:*

- *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
- *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
- *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob 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. alfa spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
  - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).
  - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.
  - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.

- ja**
- nein**

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

---



---

##### Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

- niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

- hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:

---



---

### B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: \_\_\_\_\_

#### 1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---



---

#### 2. Adäquate Umsetzung des ITT-Prinzips

Kommen in einer Studie Patienten vor, die die Studie entweder vorzeitig abgebrochen haben oder wegen Protokollverletzung ganz oder teilweise aus der Analyse ausgeschlossen wurden, so sind diese ausreichend genau zu beschreiben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen zu berücksichtigen (i. d. R. ITT-Analyse, siehe Äquivalenzstudien). Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

**unklar:** Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

**nein:** Keines der unter „ja“ genannten drei Merkmale trifft zu.

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

---

---

### 3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

*Beachte die Hinweise zu Punkt A.4!*

**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

**nein:** Es liegen Anzeichen für eine ergebnisgesteuerte Berichterstattung vor.

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

---

---

### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

*z. B.*

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

**ja**

**nein**

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

---

---

### **Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

*Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.*

**niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

**hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:

---

---