Nutzenbewertung

Gemeinsamer Bundesausschuss

von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V

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

Wirkstoff: Cerliponase alfa

Datum der Veröffentlichung: 2. Oktober 2017

Inhaltsverzeichnis

Tabellenverzeichnis 3
Abbildungsverzeichnis 5
Abkürzungsverzeichnis 6
Hintergrund 8
1 Einführung 9
2 Nutzenbewertung 11
2.1 Fragestellung 11
2.2 Zulassungsbegründende Studien 11
2.3 Liste der verwendeten Quellen 12
2.4 Studiencharakteristika 12
2.4.1 Studie 190-201/190-202 12
2.4.2 Studie 190-901 22
2.5 Endpunkte und statistische Methoden 24
2.5.1 In der Nutzenbewertung berücksichtige Endpunkte 24
2.5.2 Verzerrungspotential auf Studien- und Endpunktebene 26
2.5.3 Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz 26
2.5.4 Statistische Methoden 40
2.6 Ergebnisse zum Zusatznutzen 50
2.6.1 Charakterisierung der Studienpopulation 51
2.6.2 Mortalität 61
2.6.3 Morbidität 61
2.6.4 Lebensqualität 75
2.6.5 Subgruppenanalysen 78
2.6.6 Sicherheit 78
3 Methodische Anmerkungen und Bewertung der Unterlagen 87
3.1 Zulassungsstatus und Zulassungspopulation von Cerliponase alfa 87
3.2 Design und Methodik der Studie 190-201/190-202 88
3.3 Design und Methodik der Studie 190-901 und des historischen Vergleichs 90
3.4 Wirksamkeit 94
3.5 Lebensqualität 97
3.6 Sicherheit 98
4 Anforderungen an eine qualitätsgesicherte Anwendung 101
5 Zusammenfassung der Bewertung 102
Anhang 105
Referenzen 107

Tabellenverzeichnis

Tabelle 1: Charakterisierung der Studien 190-201/190-202 13
Tabelle 2: Charakterisierung der Intervention in den Studien 190-201/190-202 21
Tabelle 3: Charakterisierung der Studie 190-901 23
Tabelle 4: In der Nutzenbewertung berücksichtigte Endpunkte 24
Tabelle 5: Beschreibung der Endpunkte der Kategorie Mortalität 27
Tabelle 6: Beschreibung der Endpunkte der Kategorie Morbidität 27
Tabelle 7: Beschreibung der Endpunkte der Kategorie Lebensqualität 34
Tabelle 8: Beschreibung der Endpunkte der Kategorie Sicherheit 36
Tabelle 9: Studienpopulationen für den Vergleich 46
Tabelle 10: Allgemeine Angaben zur Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016) 51
Tabelle 11: Charakterisierung der Studienpopulation der Studie 190-201/190-202 (eingeschlossene Patienten [N=24], Datenschnitt: 15.10.2015) 53
Tabelle 12: Begleitmedikationen (≥ 20 %) der Studie 190-201/190-202 (eingeschlossene Patienten [N=24], Datenschnitt: 15. Oktober 2015) 56
Tabelle 13: Charakterisierung der Studienpopulation der Studie 190-901 (auswertbare Population [N=41], Datenschnitt: 01. Mai 2016) 58
Tabelle 14: Charakterisierung der 1:1-Paarbildung für den historischen Vergleich aus den Studien 190-201/190-202 vs. 190-901 61
Tabelle 15: Ergebnisse der Responderanalyse der ML-Skala der Studie 190-201 und 190-201/190-202 (ITT-Population, Dosis 300 mg) 62
Tabelle 16: Sensitivitätsanalysen zu Responderanalysen der ML-Skala der Studie 190-201 und 190-201/190-202 (Dosis 300 mg) 63
Tabelle 17: Ergebnisse zur Slope-Analyse der ML-Skala skaliert auf 48 Wochen in der Studie 190-201 und 190-201/190-202 (ITT-Population, Dosis 300 mg) 65
Tabelle 18: Sensitivitätsanalysen zur Slopeanalyse der ML-Skala skaliert auf 48 Wochen der Studie 190-201 und 190-201/190-202 (Dosis 300 mg) 65
Tabelle 19: Ergebnisse zur Veränderung der ML-Skala, MLV-Skala und MLVS-Gesamtskala zu Baseline in der Studie 190-201/190-202 (ITT-Population, Dosis 300 mg) 66
Tabelle 20: Ergebnisse der Responderanalyse der ML-/HML-Skala der Studie 190-201/190-202 (ITT-Population) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung (Dosis 300 mg, Datenschnitt: 15. Juni 2016) 68
Tabelle 21: Sensitivitätsanalysen zur primären Responderanalyse gemessen mittels ML-/HML-Skala: Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung (Dosis 300 mg, Datenschnitt 15. Juni 2016) 69
Tabelle 22: Ergebnisse der Slope-Analyse (mittlerer Abfall der ML-/HML-Skala): Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung (Dosis 300 mg, Datenschnitt 15. Juni 2016) 70
Tabelle 23: Sensitivitätsanalysen zur Slope-Analyse (mittlerer Abfall der ML-/HML-Skala): Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung (Dosis 300 mg, Datenschnitt 15. Juni 2016) 71

Tabelle 24: Ergebnisse zur Veränderung der ML-/HML-Skala zu Baseline: Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung (Dosis 300 mg, Datenschnitt 15. Juni 2016) 74

Tabelle 25: Ergebnisse zur Lebensqualität – PedsQL™-Modul „Parent Report for Toddlers“ der Studie 190-201/190-202 (ITT) (Studien-Baseline, Dosis 300 mg, Datenschnitt 03. Juni 2016) 76

Tabelle 26: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Studie 190-201/190-202 78

Tabelle 27: UE mit Inzidenz ≥ 10 % der Studie 190-201/190-202 79

Tabelle 28: UE des NCI-CTCAE-Grades 3 und höher ≥ 2 % 81

Tabelle 29: SUE mit Inzidenz ≥ 5 % 83

Tabelle 30: UEs von besonderem Interesse ≥ 10 %6) 85

Tabelle 31: Zusammenfassende Darstellung der Studienergebnisse für die einarmigen Studien 190-201/190-202 (ITT-Population [N=23], Baseline: ab 300-mg-Dosierung) 102

Tabelle 32: Zusammenfassende Darstellung der Studienergebnisse: Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung (Dosis 300 mg, Datenschnitt 15. Juni 2016) 104

Abbildungsverzeichnis

Abbildung 1: Gegenüberstellung von Klassifikation, Manifestationsalter und Genloci der NCL-Erkrankungen 9

Abbildung 2: Studienablauf der Studie 190-201 13

Abbildung 3: Unterschiede zwischen den Domänen ‚motorische und sprachliche Fähigkeiten' zwischen der ursprünglichen HML-Skala (verwendet in historischer Kontrollstudie 190-901) und der modifizierten ML-Skala 29

Abbildung 4: Zeit bis zum ersten stabilen Abfall des ML-Scores (≥ 2 Punkte) oder Score von 0 nach Kaplan-Meier der Studie 190-201/190-202 64

Abbildung 5: Zeit bis zum ersten stabilen Abfall des ML-Scores um 1 Punkt nach Kaplan-Meier für Studie 190-201/190-202 64

Abbildung 6: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML-/HML-Skala (stabile Abnahme um ≥ 2 Punkte oder Wert von Null): Cerliponase alfa (190-201/202) vs. unbehandelte Kontrolle (190-901) der ITT-Population mit 1:1-Paarbildung (Alter, ML-/HML-Wert); Datenschnitt 01.11.2016 73

Abbildung 7: Individuelle Patientendaten der HML-Skala der 29 ausgewerteten Patienten der Studie 190-901 zum natürlichen Krankheitsverlauf 105

Abbildung 8: Mittels linearer Regression modellierte Verläufe der HML-Skala der 29 ausgewerteten individuellen Patientendaten der Studie 190-901 zum natürlichen Krankheitsverlauf 106

Abkürzungsverzeichnis

CHMP	Committee for Medicinal Products for Human Use
CLN2	Ceroid-Lipofuszinose Typ 2
CSR	Clinical Study Report (Klinischer Studienbericht)
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee (Datenüberwachungskomitee)
EMA	European Medicines Agency
EPAR	European Public Assessment Report
ERT	Enzyme Replacement Therapy (Enzymersatztherapie)
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
H_0	Nullhypothese
H_1	Alternativhypothese
HML	Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg)
HR	Hazard Ratio
IAR	Infusion-Associated Reaction (Reaktion in Verbindung mit einer Infusion)
ICV	Intrazerebroventrikulär
IgE	Immunglobulin E
ISE	Integrated Summary of Efficacy
ITT	Intention-to-treat
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KI	Konfidenzintervall
LOCB	Last Observation Carried Backward
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MCID	Minimal Clinically Important Difference (Geringster klinisch relevanter Unterschied)
ML-Skala	Motor/Language (Kombinierte Skala Motorik/Sprache)
MLV-Skala	Motor/Language/Vision (Kombinierte Skala Motorik/Sprache/Sehen)
MLVS-Skala	Motor/Language/Vision/Seizures (Kombinierte Skala Motorik/Sprache/Sehen/Krampfanfälle)

MMRM	Mixed Model for Repeated Measures (Gemischtes Modell mit Messwiederholung)
MW	Mittelwert
N	Anzahl
n.b.	Nicht berichtet
NCL	Neuronale Ceroid-Lipofuszinose
PT	Preferred Term
pU	Pharmazeutischer Unternehmer
SAP	Statistischer Analyseplan
SD	Standardabweichung
SGB	Sozialgesetzbuch
SOC	Systemorganklasse
SUE	Schwerwiegende/s unerwünschte/s Ereignis/se
TPP1	Tripeptidylpeptidase 1
UE	Unerwünschte/s Ereignis/se
VerfO	Verfahrensordnung
WCMC	Weill Cornell Medical College
ZN	Zusatznutzen

Hintergrund

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

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

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

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

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Cerliponase alfa in seiner Sitzung am 26. September 2017 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 03. Juli 2017 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pharmazeutischen Unternehmers in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 02. Oktober 2017 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (http://www.g-ba.de) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1 Einführung

Neuronale Ceroid Lipofuszinose vom Typ 2 (CLN2-Erkrankung)

Die Neuronalen Ceroid-Lipofuszinosen (NCLs) sind eine Gruppe vererbter progredienter degenerativer Hirnkrankheiten. Sie gehören zu den lysosomalen Speicherkrankheiten und bilden die häufigsten neurodegenerativen Erkrankungen im Kindes- und Jugendalter. Derzeit sind 14 NCL-Formen bekannt, von denen zwei bisher nicht veröffentlicht, sondern kürzlich durch Smith et al. (CLN13) und Staropoli et al. (CLN14) bei der 13. International Conference on Neuronal Ceroid Lipofuscinosis 2012 in London vorgestellt wurden [23] [22]. Wie in Abbildung 1 deutlich wird, liegen je nach NCL-Form unterschiedliche Gendefekte vor. Charakteristisch ist bei allen Formen eine Kombination von Demenz, Visusverlust, Epilepsie und motorischem Abbau. Die Ausprägung dieser Symptome variiert, ebenso das Erkrankungsalter – von der Geburt bis in das junge Erwachsenenalter. Die meisten NCLs werden autosomal-rezessiv vererbt, nur die adulte Form CLN4 wird autosomal-dominant vererbt. Die NCLs treten mit einer geschätzten Inzidenz von 0,15:100.000 Lebendgeborenen in Portugal, 0,46:100.000 in West-Deutschland, 0,78:100.000 in England, 9,0:100.000 in Neufundland und einer geschätzten Prävalenz von 0,6–0,7 je 1.000.000 in Skandinavien auf [28].

Klinisch haben die NCL trotz ihrer Heterogenität viele Gemeinsamkeiten. Dies hat sowohl für die Diagnostik als auch für die (palliative) Behandlung Bedeutung. Dabei kann der Krankheitsbeginn zwischen Geburt und jungem Erwachsenenalter liegen. Die Reihenfolge des Auftretens der Symptome ist variabel. Erstsymptome bei Kleinkindern sind Stillstand und Rückschritte der psychomotorischen Entwicklung oder Epilepsie. Bei Schulkindern kommt es meist zunächst zu einem Visusverlust, der von einer Demenz gefolgt ist. Auf pathologischer Ebene ist den verschiedenen NCL-Formen die intrazelluläre Speicherung von autofluoreszierendem Material mit charakteristischer Ultrastruktur in fast allen Geweben gemeinsam. Im Gehirn geht der Speicherprozess mit dem Absterben von Neuronen einher [23].

Gendefekt	Klinischer Typ	Manifestationsalter	Genlocus
CLN1	Infantile NCL (INCL)	0–2 (bis 38) Jahre	1p32
CLN2	Spätinfantile NCL (LINCL)	2–8 Jahre	11p15
CLN3	Juvenile NCL (JNCL)	4–10 Jahre	16p12
CLN4	Adulte NCL (ANCL)	Um 30 (11–55) Jahre	Unbekannt
CLN5	Finnische spätinfantile NCL	4–7 Jahre	13q21.1–32
CLN6	Indisch-iberische NCL	1–8 Jahre	15q21–23
CLN7	Türkische NCL	1–6 Jahre	4q28.1–28.2
CLN8	Nordische Epilepsie	5–10 Jahre	8p23
CLN9	CLN9-defiziente NCL	Um 4 Jahre	Unbekannt
CTSD (CLN10)	Kongenitale NCL (CNCL)	0–6 Jahre	11p15.5

Abbildung 1: Gegenüberstellung von Klassifikation, Manifestationsalter und Genloci der NCL-Erkrankungen [20]

Bei dem hier vorliegenden Anwendungsgebiet ‚Neuronale Ceroid-Lipofuszinose vom Typ 2 (im Folgenden als „CLN2-Erkrankung“ bezeichnet) liegt ein Mangel der lysosomalen Serin-Protease Tripeptidyl Peptidase 1 (TPP1) vor [19]. Wie eingangs beschrieben, wird die Erkrankung autosomal-rezessiv vererbt, durch die Mutationen des CLN2-Gens verursacht und manifestiert sich

vorwiegend als spät-infantile Form der NCL [15] [19]. Die CLN2-Erkrankung tritt symptomatisch im Alter zwischen zwei bis vier Jahren mit dem Auftreten von Krampfanfällen, dem Verlust der Sprache, der motorischen und kognitiven Funktionen sowie – etwas später – dem Verlust der Sehfähigkeit auf. Die Krankheit schreitet nach Auftreten der ersten Symptome rapide fort und führt zu einem vollständigen psychomotorischen Funktionsverlust im Alter von etwa 6 Jahren. Pathologisch tritt ein massiver neuronaler Zelltod ein [28]. Die Lebenserwartung beträgt 10 bis 15 Jahre [23].

Cerliponase alfa

Cerliponase alfa (BMN 190) ist ein Wirkstoff aus der Gruppe der rekombinanten Enzyme, der zur Behandlung der spätinfantilen Neuronalen Ceroid-Lipofuszinose eingesetzt wird. Cerliponase alfa ist eine rekombinante Tripeptidylpeptidase 1 (rhTPP1), die mit biotechnologischen Methoden hergestellt wird. Cerliponase alfa ist die erste Enzymersatztherapie zur Behandlung von Patienten mit einer CLN2-Erkrankung, auch als Tripeptidyl-Peptidase-1-(TPP1)-Mangel bezeichnet. Cerliponase alfa wird mittels intrazerebroventrikulärer Infusion (ICV) appliziert [1].

Der Wirkstoff Cerliponase alfa (Brineura®) ist eine Kopie von TPP1 und dient als Ersatz für das fehlende Enzym. Brineura® wird direkt in das Gehirn mittels ICV infundiert. Vor der ersten Infusion muss sich der Patient einer Operation zur Implantation einer Vorrichtung (ICV-Port) unterziehen, die von außen durch den Schädel zum Liquorraum im Gehirn führt, wo das Arzneimittel verabreicht wird. Die Infusionen werden einmal alle zwei Wochen von medizinischem Fachpersonal mit Kenntnissen in der Verabreichung von Arzneimitteln im Gehirn verabreicht [17].

2 Nutzenbewertung

2.1 Fragestellung

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

- Cerliponase alfa ist angezeigt zur Behandlung der neuronalen Ceroid-Lipofuszinose (NCL) Typ 2, auch als Tripeptidyl-Peptidase-1-(TPP1)-Mangel bezeichnet.

2.2 Zulassungsbegründende Studien

Pivotale (Phase I/II) Studie

- Studie 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“. Bei der Studie 190-201 handelt es sich um eine nicht-kontrollierte, einarmige und offene Phase-1/2-Dosis-Eskalations-Studie mit dem Ziel, die Sicherheit, Akzeptanz, Pharmakokinetik und Wirksamkeit von Cerliponase alfa bei der spät-infantilen CLN2-Erkrankung zu untersuchen.

Supportive Studie(n)

- Studie 190-202: „A Multicenter, Multinational, Extension Study to Evaluate the Long-Term Efficacy and Safety of BMN 190 in Patients with CLN2 Disease“. Bei der Studie 190-202 handelt es sich um eine multizentrische, multinationale Extensionsstudie der Studie 190-201, um die langfristige Wirksamkeit und Sicherheit von Cerliponase alfa bei der spät-infantilen CLN2-Erkrankung zu untersuchen.
- Studie 190-901: „Natural History of Late-Infantile CLN2 Disease: Quantitative Assessment of Disease Characteristics, Rate of Progression, and Magnetic Resonance Imaging Findings.“ Bei der Studie 190-901 handelt es sich um eine historische Kohorte für Patienten mit einer spät-infantilen CLN2-Erkrankung mit dem Ziel, die Charakteristik der Erkrankung, die Rate der Progression und bildgebende Analysen mit der Magnetresonanztomographie zu bewerten. Die aktualisierte Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901) dient als historische Kontrolle der Studien 190-201/190-202. Bei dieser Analyse wurden spezifische Auswahlkriterien eingesetzt, um eine Patientenpopulation zu erhalten, die den wesentlichen Ein- und Ausschlusskriterien der Studie 190-201/190-202 entsprach.
- Cerliponase alfa wurde unter „außergewöhnlichen Umständen“ (exceptional circumstances) zugelassen, da es aufgrund der Seltenheit der Krankheit nicht möglich war, vollständige Informationen über Cerliponase alfa (Brineura®) zu erhalten [17]. Die Zulassung unter „außergewöhnlichen Umständen“ ist mit der Auflage weiterer Studienergebnisse verknüpft: Eine nicht-interventionelle post-authorisation Sicherheitsstudie (190-501) zur Beurteilung der Langzeitsicherheit einschließlich dem Auftreten von Überempfindlichkeitsreaktionen und

Anaphylaxien auf Basis einer adäquaten Datenquelle, die sich aus einer Registrierung von Patienten mit CLN2 ergibt (Registerstudie). Ergebnisse sollen jährlich für die jährliche Neubewertung vorlegt werden. Eine nicht-interventionelle post-authorisation Wirksamkeitsstudie (190-203) zur Beurteilung der Krankheitsprogression anhand der motorischen klinischen Skala sowie der Sicherheit und Verträglichkeit bei Patienten unter 2 Jahren (N≥5). Ergebnisse sind im Jahr 2020 zu erwarten [18].

2.3 Liste der verwendeten Quellen

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

- Herstellerdossier zu Cerliponase alfa
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [18]
- Studienbericht inkl. Studienprotokollen und statistischem Analyseplan der Studie 190-201
  - CSR vom 27. April 2016 (Studie beendet: 30. November 2015 [10,11,12]
- Studienbericht inkl. Studienprotokollen und statistischem Analyseplan der Studie 190-202
  - CSR: 15. Oktober 2015 [6,7,12]
  - CSR: 03. Juni 2016 [8]
- Studienbericht (inkl. statistischem Analyseplan) der Studie 190-901 *(Hinweis: es existieren keine Studienprotokolle)* [9,14]
- Studienbericht und statistischer Analyseplan der ISE-Analyse *(für den historischen Vergleich)*
  - CSR: 30. November 2015 [3,4]
  - CSR: 15. Juni 2016 [5]
  - CSR: 01. November 2016 [13]
- Validierungsstudie zum primären Endpunkt ‚ML-/HML-Skala' [31]

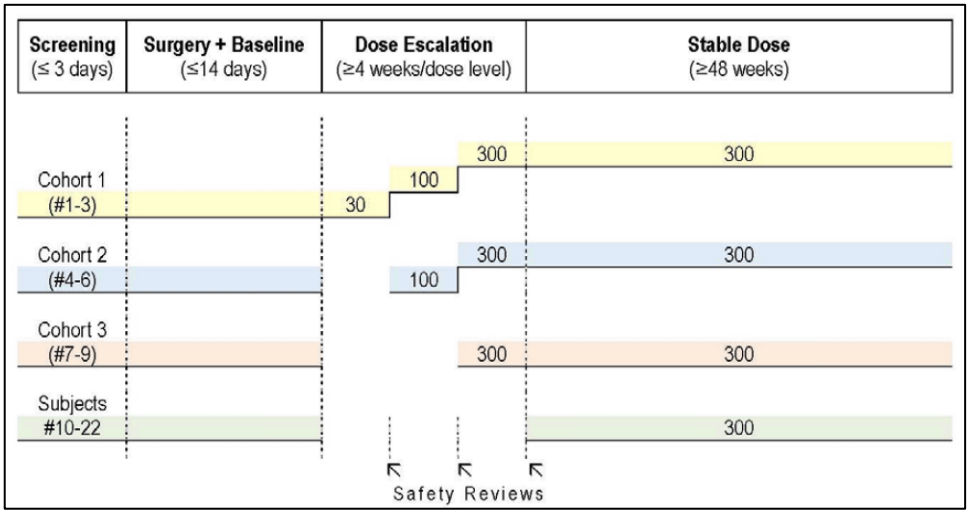
2.4 Studiencharakteristika

Die Unterlagen zum Nachweis des Ausmaßes des Zusatznutzens für Cerliponase alfa basieren auf der pivotalen (PhaseI/II) Zulassungsstudie 190-201, der Extensionsstudie 190-202 sowie der supportiven Studie 190-901. Die Studien und die Intervention werden in den Tabellen 1 bis 3 charakterisiert.

2.4.1 Studie 190-201/190-202

Das Studiendesign, die Patientenpopulation und die Interventionen der Open-Label-Dosis-Eskalationsstudie (190-201) sowie der noch andauernden langfristigen Open-Label-Verlängerungsstudie (190-202) werden in den Tabellen 1 bis 2 charakterisiert.

Tabelle 1: Charakterisierung der Studien 190-201/190-202

Charakteris-tikum	Beschreibung
Design	**Studie 190-201** Bei der Studie 190-201 handelt es sich um eine ‚first-in-human Phase 1/2-Studie' - eine multizentrische, offene, einarmige Interventionsstudie, die sich in zwei Phasen gliedert. • Phase 1: Dosis-Eskalationsphase (≥ 4 Wochen je Dosis-Level, d. h. Patienten erhielten initial Dosen von 30, 100 bzw. 300 mg Cerliponase alfa ICV alle zwei Wochen über jeweils mindestens vier Wochen) • Phase 2: stabile Dosierungsphase (für mindestens 48 Wochen 300 mg Cerliponase alfa) Studienpopulation: Patienten zwischen drei und 15 Jahren mit einer milden bis moderaten CLN2-Erkrankung unter der Annahme eines Scores von 3–6 auf der ML-Skala und einem ML-Score von mindestens 1 in jeder der beiden Domänen sowie einer bestätigten Genmutation des CLN2/TPP1-Gens. Applikationsform: Die Patienten erhalten den Wirkstoff Cerliponase alfa intrazerebroventrikulär (ICV) alle zwei Wochen. Ziel der Studie: Sicherheit, Akzeptanz, Pharmakokinetik und Wirksamkeit Gemäß dem finalen SAP (Version 3, 25. März 2016) der Studie 190-201/190-202 war für die primäre Auswertung ein Vergleich mit einer historischen Kontrollgruppe vorgesehen. Dieser wurde auf Empfehlung der Zulassungsbehörden im Statistischen Analyseplan der „Integrated Summary of Efficacy" (SAP ISE) spezifiziert: „Auf Empfehlung der Zulassungsbehörden wurde im ISE SAP eine Responder-Analyse nach 1:1-Paarbildung der Teilnehmer der Studie 190-201 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) als primärer Endpunkt festgelegt" [2]. Die aktualisierte Analyse erfolgte auf Basis des DEM-CHILD-Datensatzes ‚Supplementary Report der Studie 190-901' (nähere Informationen: 2.4.2). Aufbau der Studie: Die Studie umfasst eine Screeningphase (≤ 3 Tage), Operationsphase (≤ 14 Tage), eine Dosis-Eskalationsphase (≥ 4 Wochen je Dosis-Level) und eine stabile Dosisphase 300 mg (≥ 48 Wochen) (siehe nachfolgende Abbildung)  *Abbildung 2: Studienablauf der Studie 190-201 [10]* Screeningphase (≤ 3 Tage) • Überprüfung der Einschlusskriterien • Einverständniserklärung

Charakteris-tikum	Beschreibung
	Operationsphase (≤ 14 Tage) • Chirurgische Implantierung eines ICV-Reservoirs und einer Kanüle in den lateralen Ventrikel der rechten Hemisphäre • Die erste Infusion des Wirkstoffs kann sowohl während als auch nach dem chirurgischen Eingriff gegeben werden, allerdings nicht 14 Tage nach der Operation. • Erfassung der Baseline-Daten (klinische Scores und klinische Laborparameter), wird nicht länger als 2 Tage vor der Gabe der ersten Infusion erhoben Dosis-Eskalationsphase (≥ 4 Wochen je Dosis-Level) • Neun Patienten werden in der Dosis-Eskalationsphase in drei Kohorten unterteilt (siehe Abbildung 2 [10]): o Kohorte 1: Patienten 1 bis 3 (starten bei 30 mg Cerliponase alfa) o Kohorte 2: Patienten 4 bis 6 (starten bei 100 mg Cerliponase alfa) o Kohorte 3: Patienten 7 bis 9 (starten bei 300 mg Cerliponase alfa) • Der Zeitraum der Dosis-Eskalationsphase umfasst für alle 3 Kohorten mindestens zwei Gaben/Verabreichungen je geplantem Dosislevel. • Das Datenkontrollkomitee (DMC) entschied auf Basis der Sicherheitsdaten über die Einschreibung der Patienten in das nächst höhere Dosisschema, nachdem der letzte Patient in der jeweiligen Kohorte die zweite Dosis auf dem Dosis-Level beendet hatte. Die jeweils nächst höhere Dosis wurde nur verabreicht, wenn die Patienten die niedrigere Dosis vertragen hatten (keine unakzeptable Toxizität); hatten mindestens 2 Patienten eine nicht-akzeptable toxische Reaktion aufgrund der Medikation, wurde das Dosis-Level als nicht empfehlenswert eingestuft; die Eskalation der Kohorten-Start-Dosis wird nicht empfohlen, wenn mindestens ein Patient eine nicht-akzeptable toxische Reaktion aufgrund der Medikation erlebt. stabile Dosisphase 300 mg (≥ 48 Wochen) • Die Einschreibung der Patienten 10 bis 22 erfolgt, nachdem die Patienten eins bis neun mindestens zwei Dosen Cerliponase alfa 200 mg erhalten haben und auf Basis der Entscheidung durch das DMC. • Nach Beendigung der Dosis-Eskalationsphase gehen ebenfalls alle neun Patienten der Dosis-Eskalationsphase in die stabile Dosisphase über, wenn sie für 4 Wochen die Dosis von 300 mg oder die höchst-tolerierbare Dosis toleriert haben. • Dosierung: 300 mg Cerliponase alfa oder die höchsttolerierbare Dosis **Studie 190-202** Bei der Studie 190-202 handelt es sich um eine multizentrische, offene, einarmige Extensionsstudie der Studie 190-201. Wie oben bereits beschrieben, wurde „auf Empfehlung der Zulassungsbehörden im ISE SAP eine Responder-Analyse nach 1:1-Paarbildung der Teilnehmer der Studie 190-201 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) als primärer Endpunkt festgelegt“ [2]. Die aktualisierte Analyse erfolgte auf Basis des DEM-CHILD-Datensatzes ‚Supplementary Report der Studie 190-901‘ (nähere Informationen: 2.4.2). In Woche 47 der Studie 190-201 begann eine zweiwöchige Screeningphase für die sich anschließende Extensionsstudie 190-202. Die erste Infusionsgabe (300 mg Cerliponase alfa/ alle 2 Wochen) erfolgte in Woche 49. Studienpopulation: Patienten, die die Studie 190-201 abgeschlossen hatten (48 Wochen) Ziel der Studie: Langzeit-Sicherheit und Langzeit-Wirksamkeit Studiendauer: 239 Wochen (geplantes Studienende: 15. Dezember 2020)

Charakteris-tikum	Beschreibung
Population	**Studie 190-201** Wesentliche Einschlusskriterien: • Gesicherte CLN2-Diagnose (mittels TPP1-Enzymaktivität im Blut und CLN2-Genanalyse) • leichte bis mittelschwere Erkrankung (Gesamtscore von 3–6 auf der ML-Skala und ein ML-Score von mindestens 1 in jeder der beiden Domänen) • schriftliche Einverständniserklärung durch ein Elternteil oder einen gesetzlichen Vertreter und, falls möglich, Zustimmung des Patienten • Fähigkeit, an den protokollspezifischen Untersuchungen teilzunehmen • Krampfanfälle und begleitende Anti-Epileptika aus Sicht des Prüfarztes stabil Wesentliche Ausschlusskriterien: • Alter < 3 Jahre oder ≥ 16 Jahre zum Zeitpunkt des Studieneinschlusses • andere neurologische Erbkrankheit wie z. B. andere CLN oder nicht mit CLN2 in Zusammenhang stehende Krampfanfälle (Einschluss von Patienten mit Fieberkrämpfen möglich) • andere neurologische Erkrankung vor Studieneinschluss, die zum Verlust kognitiver Fähigkeiten geführt haben könnte (z. B. Trauma, Meningitis, Blutung) • Unterstützende Beatmungsmaßnahmen ausgenommen nicht invasive Maßnahmen in der Nacht • Vorausgegangene Stammzell- oder Gentherapie oder ERT zur Behandlung der CLN2 • Kontraindikationen für einen neurochirurgischen Eingriff (z. B. angeborener Herzfehler, schwere respiratorische Insuffizienz oder Gerinnungsstörung) • Kontraindikationen für eine Kernspintomographie • Generalisierter, konvulsiver Status epilepticus innerhalb von vier Wochen vor der ersten geplanten Dosis (klinische und nicht nur EEG-basierte Diagnose) • Schwere Infektion (z. B. Pneumonie, Pyelonephritis oder Meningitis) vor der ersten geplanten Dosis (eventuell späterer Einschluss) • Risiko ICV-bedingter Komplikationen wie z. B. Hydrozephalus oder Ventrikel-Shunt • Überempfindlichkeit gegenüber einem der Inhaltsstoffe des Prüfpräparats • Verabreichung eines anderen Prüfpräparates innerhalb von 30 Tagen vor der ersten Infusion der Studienmedikation oder geplante Gabe eines anderen Prüfpräparates im Verlauf der Studie **Studie 190-202** Wesentliche Einschlusskriterien: • Eingeschlossen wurden alle Patienten, die die 48 Wochen der Studie 190-201 beendet hatten. • schriftliche Einverständniserklärung durch ein Elternteil oder einen gesetzlichen Vertreter und, falls möglich, Zustimmung des Patienten Wesentliche Ausschlusskriterien: • Abfall des ML-Scores um ≥ 3 Punkte während der Studie 190-201 (von Studienbeginn bis Studienende) • ML-Score von 0 Punkten • Anwendung anderer Prüfparate oder investigativer Medizinprodukte innerhalb von 30 Tagen vor oder während der Studie • Begleiterkrankungen oder andere Bedingungen, die nach Ansicht des Prüfarztes gegen eine Teilnahme sprechen oder ein Risiko darstellen • Möglicherweise schlechte Compliance oder hohes Risiko eines Studienabbruchs aus Sicht des Prüfarztes

<table>
<tr><th>Charakteris-tikum</th><th>Beschreibung</th></tr>
<tr><td>Intervention und Zahl der Patienten</td><td>

Studie 190-201/190-202

	Studie 190-201[1)] N (%)	**Studie 190-202[2)] N (%)**
Eingeschlossene Patienten	24 (100) 23 (ITT-Population[3)])	23 (96)[3)]

[1)] Studie abgeschlossen am 30. November 2015. Alle Patienten der Studie 190-201 (N=23, ITT) hatten zu diesem Datenschnitt die 48 Wochen der Behandlung mit 300 mg Cerliponase abgeschlossen.
[2)] Datenschnitt: 03. Juni 2016. Bis zu diesem Zeitpunkt waren 23 (96 %) der ursprünglich eingeschlossenen 24 Patienten über mindestens 72 Wochen in fester Dosierung (300 mg Cerliponase alfa) behandelt worden.
[3)] Ein Patient zog seine Einverständniserklärung nach einer Infusion zurück.

Studie 190-901

	Studie 190-901 N
Original-Datenbank[1)]	79
DEM-CHILD-Datenbank[2)]	74

[1)] Original-Analyse (30.06.2015): In die Original-Datenbank wurden Patienten aus zwei großen Kohorten in Hamburg (Längsschnittdaten n=30, davon 29 ausgewertet) bzw. in New York (Querschnittsdaten n=49; WCMC [Weill Cornell Medical College]) eingeschlossen (nähere Informationen siehe Kapitel 2.4.2).
[2)] Supplemental Report (01.05.2016): DEM-CHILD-Register stellt eine Erweiterung des ursprünglichen Datensatzes (Original-Datenbank) aus Hamburg dar. Zum Zeitpunkt der Analyse (Februar 2015) lagen Registerdaten von CLN2-Patienten aus zwei Zentren vor (Hamburg, n=63, Verona, n=11). Es handelt sich um eine europaweite Datenbank mit Sitz in Hamburg (nähere Informationen siehe Kapitel 2.4.2).

Historischer Vergleich: Studie 190-201/190-202 vs. 190-901[5)]

	Studie 190-201/190-202 N=23[3)]	**Studie 190-901[1)] N=41[2)]**
1:1 matching[4)6)]	22	22
1:1 matching[4)7)]	21	21

[1)] Von den insgesamt 74 Patienten im DEM-Child-Register stammten 63 aus dem Zentrum in Hamburg und 11 aus dem Zentrum in Verona.
[2)] Die aktualisierte Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901) dient als historische Kontrolle der Studien 190-201/190-202. Bei dieser Analyse wurden spezifische Auswahlkriterien eingesetzt, um eine Patientenpopulation zu erhalten, die den wesentlichen Ein- und Ausschlusskriterien der Studie 190-201 entsprach. Nach Abgleich dieser Kriterien und Ausschluss eines Zwillings (nähere Informationen 2.4.2) verblieben 41 von 74 Patienten (55 %) als auswertbare Patienten in der Analyse und erfüllten die Eignungskriterien für den historischen Vergleich mit den Patienten der Studie 190-201/190-202.
[3)] ITT-Population: Alle Patienten, die die Studienmedikation erhalten haben, unabhängig von der Dosis, und bei denen zumindest ein Wirksamkeitsergebnis vorlag. Ein Patient der ursprünglich 24 eingeschlossenen Patienten 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 Dosis-Eskalationsphase erhalten hatte und der Abbruch nicht mit dem Behandlungsergebnis in Zusammenhang stand.
[4)] Matching-Kriterien: ML-Wert (Studie 190-201/202) und HML-Wert (Studie 190-901) zu Baseline stimmten exakt überein und der Altersunterschied (Baseline) war so gering wie möglich (nicht mehr als 12 Monate).
[5)] vom pU vorgelegte Datenschnitte für den Vergleich: 30. November 2015; 15. Juni 2016; 01. November 2016
[6)] Datenschnitt: 15. Juni 2016 (umfasst für die Studie 190-201/190-202 den Datenschnitt vom 03. Juni 2016 und für die Studie 190-901 den Datenschnitt vom 11. August 2016 [zeitliche Diskrepanz bleibt nach Sichtung der eingereichten Unterlagen unklar]).
[7)] Datenschnitt: 01. November 2016. Bis zu diesem Zeitpunkt waren alle Patienten der Studie 190-201/190-202 über mindestens 96 Wochen mit Cerliponase alfa in fester Dosierung (300 mg Cerliponase alfa) behandelt worden.

</td></tr>
</table>

<table>
<tr><th>Charakteris-tikum</th><th>Beschreibung</th></tr>
<tr><td>Ort und Zeitraum der Durch-führung</td><td>
<table>
<tr><th>Studie 190-201</th><th>Studie 190-202</th></tr>
<tr><td>Studienbeginn (erster Patient eingeschlossen)[3]: 13.09.2013
Studienende (letzte Visite): 30.11.2015
Interims-Data-Cutoffs: September 2014 und Januar 2015
Erstellung Studienbericht: 27.04.2016</td><td>Studienbeginn (erster Patient eingeschlossen): 02.02.2015
Studienende (letzte Visite): 15.12.2020
Interims-Data-Cutoff 1[1]: 15.10.2015
Interims-Data-Cutoff 2[2]: 03.06.2016
Erstellung Studienbericht: 09.05.2016</td></tr>
</table>
[1] Nicht a priori definiert.
[2] Nicht a priori definiert. Wurde auf Ersuchen der FDA durchgeführt.
[3] Der erste Patient wurde unter Amendement 2 des Studienprotokolls vom 26. Juli 2013 eingeschlossen.

<u>Ort der Durchführung (für Studie 190-201/190-202):</u>
Die Studie wurde in 5 klinischen Zentren in 4 Ländern durchgeführt: 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.
</td></tr>
<tr><td>Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte gemäß Studien-protokoll</td><td>
Studie 190-201

<u>Primäres Zielkriterium:</u>
• Sicherheit und Verträglichkeit von Cerliponase alfa bei Patienten mit CLN2
• Wirksamkeit von Cerliponase alfa nach 12 Monaten verglichen mit einer unbehandelten historischen Kontrollgruppe (gemessen mittels einer CLN2-spezifischen Bewertungsskala, ML-Skala)

<u>Sekundäre Zielkriterien:</u>
• Effekt der Behandlung auf die Gehirnatrophie mittels MRT verglichen mit einer unbehandelten historischen Kontrollgruppe nach 12 Monaten
• Charakterisierung der Pharmakokinetik nach Einmal- und Mehrfachgabe (CSF und Plasma)
• Immunogenizität (CSF und Plasma)

<u>Explorative Endpunkte:</u>
• Effekt von Cerliponase alfa auf krankheitsspezifische CSF- und Plasma-Biomarker
• Dokumentation altersgerechter Entwicklungsstufen
• Krankheitsspezifische Lebensqualität (PedsQL™, “CLN2 Disease-based Quality of Life”, modifiziert auf Basis des PedsQL)

<u>Sicherheit:</u>
• Die Sicherheit wird durch die Untersuchung von UEs, SUEs, UE von besonderem Interesse, klinische Laborergebnisse (einschließlich Chemie, Hämatologie, Urinanalyse und CSF (Liquor cerebrospinalis), Vitalzeichen, EKG, EEG und Immunogenität untersucht.

<u>UE von besonderem Interesse (a priori definiert):</u>
• Überempfindlichkeitsreaktionen
• UEs, die mit der Infusion in Zusammenhang stehen („Infusion-associated reactions“, IAR)
• Unerwarteter, nicht anders erklärbarer und rascher Abfall des CLN2-Krankheits-scores
• ZNS betreffende UEs (Status epilepticus, Hydrozephalus, Meningitis)
</td></tr>
</table>

Charakteristikum	Beschreibung
	• UEs, die mit dem ICV-Zugang in Zusammenhang standen (*Hinweis: wurde in der Studie 190-201 als UE von besonderem Interesse berichtet, allerdings wurde dies nicht a priori für die Studie 190-201 definiert, sondern erst für die Studie 190-202)* **Studie 190-202** Primäres Zielkriterium: • Langzeitsicherheit von Cerliponase alfa 300 mg alle zwei Wochen bei Patienten mit CLN2 • Änderung des Summenscores aus der adaptierten, motorischen und Sprachdomäne (ML-Skala) unter Behandlung mit Cerliponase alfa Sekundäre Zielkriterien: • Änderung quantitativer Parameter im MRT • Änderung des krankheitsspezifischen CLN2-Gesamtscores • QoL unter Langzeitbehandlung mit Cerliponase alfa (PedsQL™, CLN2-spezifischer Fragebogen, EQ-5D-5L) Explorative Endpunkte: • Erreichen altersgemäßer Entwicklungsschritte („Meilensteine“) unter Langzeitbehandlung mit Cerliponase alfa • Einfluss der Behandlung auf krankheitsspezifische Biomarker in der CSF oder im Plasma Sicherheit: • Die Sicherheit wird durch die Untersuchung von UEs, SUEs, UE von besonderem Interesse, klinische Laborergebnisse (einschließlich Chemie, Hämatologie, Urinanalyse und CSF (Liquor cerebrospinalis), Vitalzeichen, EKG, EEG und Immunogenität untersucht. UE von besonderem Interesse (a priori definiert): • Überempfindlichkeitsreaktionen • UEs, die mit der Infusion in Zusammenhang stehen („Infusion-associated reactions“, IAR) • Unerwarteter, nicht anders erklärbarer und rascher Abfall des CLN2-Krankheitsscores • ZNS betreffende UEs (Status epilepticus, Hydrozephalus, Meningitis) • UEs, die mit dem ICV-Zugang in Zusammenhang standen (*Hinweis: wurde in der Studie 190-201 als UE von besonderem Interesse berichtet, allerdings wurde dies nicht a priori für die Studie 190-201 definiert, sondern erst für die Studie 190-202)*
Subgruppenanalysen	**Studie 190-201/190-202** A priori geplant In den Studienprotokollen der Studie 190-201 und 190-202 waren keine Subgruppen-Analysen festgelegt. Gemäß Dossier wurde allerdings mit Amendement 2 des Protokolls vom 07.03.2013 (gültig zum Zeitpunkt, als der erste Patient eingeschlossen 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 vom 25. März 2016) konnten folgende Subgruppen-Merkmale hinsichtlich einer möglichen Effektmodifikation (Wirksamkeit) untersucht werden: • häufiger Genotyp (insbesondere c.622C>T- und c.509-1G>C-Mutationen) • demographische Merkmale

Charakteris-tikum	Beschreibung
	• Studienzentrum oder geographische Region • verschiedene Behandlungsintervalle (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. Gemäß Modul 4 waren Homogenitäts- oder Interaktionstests nicht vorgesehen. Die folgenden weiteren Analysen ‚Veränderung zu 300-mg-Baseline' wurden im Studienbericht berichtet, waren aber nicht a priori im Studienprotokoll oder SAP geplant [11]. • 6-Punkt ML-Skala • 9-Punkt MLV-Skala (ML-Skala plus Sehvermögen) • 12-Punkt MLVS-Skala (ML-Skala plus Sehvermögen und Krampfanfälle) **Studie 190-201/190-202 vs. 190-901 (ISE SAP, Version 3 vom 25. April 2016)** <u>A priori geplant</u> • Verschiedene Behandlungsintervalle (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). <u>Post-hoc für das Dossier ausgewertete Subgruppen (nicht im Studienbericht)</u> • Abfall des ML-Scores über 48 Wochen getrennt nach Studienzentren • Die folgenden weiteren Analysen ‚Veränderung zu 300-mg-Baseline' wurden post-hoc im Studienbericht berichtet, waren aber nicht a priori im Studienprotokoll oder SAP geplant [11]. • 6-Punkt ML-Skala • 9-Punkt MLV-Skala (ML plus Sehvermögen) • 12-Punkt MLVS-Skala (MLV plus Krampfanfälle)

Abkürzungen: CLN: Ceroid-Lipofuszinose; CLN2: Ceroid-Lipofuszinose Typ 2; CSF: Cerebrospinal fluid (Liquor cerebrospinalis); DMC: Data Monitoring Committee (Datenüberwachungskomitee); EEG: Elektroenzephalogramm; ERT: Enzymersatztherapie; IAR: Infusion-Associated Reaction (Reaktion in Verbindung mit einer Infusion); ICV: Intrazerebroventrikulär; HML: Hamburg Motor/ Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg); mg: Milligramm; ML: Motor/Language (Kombinierte Skala Motorik/Sprache); SAP: statistischer Analyseplan; SUE: schwerwiegende UE; TPP1: Tripeptidylpeptidase 1; UE: unerwünschte Ereignisse, ZNS: zentrales Nervensystem

Studie 190-201

Insgesamt wurden fünf Änderungen am Studienprotokoll (Original vom 07. März 2013) der Studie 190-201 vorgenommen. Das finale Protokoll (Version 5) wurde am 24. September 2014 angefertigt. Die wichtigsten Änderungen sind nachfolgend differenziert nach Amendement dargestellt.

<u>Amendement 1 am 10. Mai 2013</u>

- Da Überempfindlichkeitsreaktionen mit ERT verbunden sein können, ist die Gabe von Antihistaminen und zusätzlicher Antipyretika (ca. 30 min vor der Infusion von Cerliponase alfa) möglich.
- Ausschlusskriterium „Patienten im Alter von 16 Jahren oder älter" wurde, entsprechend den Empfehlungen des Europäischen Ethikausschusses, hinzugefügt.
- Nachbeobachtungszeit wurde von vier Wochen auf sechs Monaten verlängert.

Amendement 2 am 26. Juli 2013

- Studieneinschluss der Patienten in die Studie 190-201 erfolgte unter Amendement 2
- Es wurde festgelegt, dass der SAP in seiner aktuellen Version 3 (Stand: 26. März 2016) gilt, sollte es Diskrepanzen zwischen der Beschreibung der statistischen Analyse im Protokoll und dem finalen SAP geben.
- Notwendigkeit einer regelmäßigen Überprüfung der Funktion des ICV-Zugangs

Amendement 3 am 07. Juli 2014

- Hinzufügen von Videoaufnahmen aller Erhebungen der ML-Skala zur Überprüfung und Ergänzung der klinischen Daten
- 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.
- Der Prüfarzt sollte bewerten, ob ein UE mit dem ICV-Zugang in Zusammenhang stand oder nicht, um eine bessere Einschätzung dieser UEs zu ermöglichen
- Klarstellung, dass alle Screening-Untersuchungen innerhalb von ≤ 3 Tagen vor der Operation (ICV) abgeschlossen sein müssen
- Zusätzliche Bestimmung des gesamten und Cerliponase-alfa-spezifischen Immunglobulin E (IgE) zu Baseline, d. h. vor der ersten Studienmedikation, um die Ergebnisse bei Auftreten einer Überempfindlichkeitsreaktion besser bewerten zu können

Amendement 5 am 24. September 2014

- 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.

Studie 190-202

Insgesamt wurden drei Änderungen am Studienprotokoll (Original vom 03. Oktober 2014) der Studie 190-202 vorgenommen. Das finale Protokoll (Version 3) für die Zwischenanalyse (15. Oktober 2015) wurde am 26. Februar 2016 angefertigt. Darüber hinaus gibt es noch zwei weitere Protokolle (Amendement 4: 17. März 2017; Amendement 5: 05. Mai 2017), die nach dem aktuellen Datenschnitt vom 03. Juni 2016 angefertigt wurden. Die wichtigsten Änderungen sind nachfolgend differenziert nach Amendement dargestellt.

Amendement 3 am 26. Februar 2016

- Keine CSF-Probennahme während der Nachbeobachtung; stattdessen Überprüfung der Immunogenizität im Serum

Amendement 4 am 17. März 2017

- Kardiovaskuläre UEs und EKG-UEs wurden als UE von besonderem Interesse hinzugefügt.
- Im Falle dessen, dass ein ICV-Port ersetzt werden musste, erfolgt die nächste Infusionsgabe nach mindestens 14 Tagen und nicht später als 28 Tage nach dem Ersatzeingriff.

Studie 190-901 und Integrated Summary of Efficacy (ISE)

Sowohl für den Natural History Report (Studie 190-901) als auch für das ISE (190-201/190-202 vs. 190-901) liegen keine Protokolle vor.

Tabelle 2: Charakterisierung der Intervention in den Studien 190-201/190-202

Intervention	Kontrolle
Studie 190-201 Name des Wirkstoffs: • Cerliponase alfa Dosis: • **Teil 1:** Dosis-Eskalation mit Startdosis von 30 mg (100 mg, 300 mg) alle 2 Wochen für mindestens 4 Wochen • **Teil 2:** Stabile Dosis mit 300 mg Cerliponase alfa alle 2 Woche für mindestens 48 Wochen Darreichungsform: • Intrazerebroventrikuläre Infusion (ICV); hierfür wurde den Patienten ein ICV-Reservoir und eine ICV-Kanüle chirurgisch implantiert. • Cerliponase alfa wurde alle zwei Wochen jeweils morgens, nachdem die Patienten mindestens zwei Stunden gefastet hatten, ICV verabreicht. Darreichungsbedingungen: • Eine Dosiseskalation (30 auf 100 mg bzw. 100 auf 300 mg) fand erst statt, nachdem Auswertungen bei der mit der jeweils niedrigeren Dosis behandelten Kohorte vorlagen; das unabhängige Datenkontrollkomitee (DMC) entschied auf Basis der Sicherheitsdaten über die Dosiseskalationen. • Das Prüfpräparat durfte nur durch qualifiziertes Krankenhauspersonal verabreicht werden. Die Patienten erhielten die Studienmedikation bis zum Auftreten einer Krankheitsprogression, inakzeptabler Toxizität oder der Rücknahme der Einverständniserklärung. Hinweise/ Safety Monitoring: • Bei Dosen unter 300 mg wurde die Lösung auf 10 ml verdünnt. • Die Flussrate der Infusion betrug 2,5 ml/Stunde bei einer Infusionsdauer von etwa vier Stunden. • Wenn die Dosis aus Sicherheitsgründen oder anderen Gründen gestoppt werden musste, konnte diese mit der gleichen Flussrate neu gestartet und abgeschlossen werden, vorausgesetzt, die Gesamtdosis wurde innerhalb von 10 Stunden nach der Vorbereitung der Dosisspritzen verabreicht. • 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). • Bei Auftreten schwerer Überempfindlichkeitsreaktionen wie Angioödem oder Stridor sollte die Infusion abgebrochen werden. • Den Patienten wurde etwa 30 Minuten vor den Infusionen ein Anti-Histaminikum (und wenn angezeigt ein Anti-Pyretikum) in altersgerechter Dosierung verabreicht. • Ein Sedativum (entsprechend dem Alter der Kinder) konnte im Ermessen der Prüfärzte vor den Infusionen eingesetzt werden.	Einarmige Studie, kein Kontrollarm

Studie 190-202 Name des Wirkstoffs: • Cerliponase alfa Dosis: • Stabile Dosis mit 300 mg alle 2 Woche für 239 Wochen Darreichungsform: • intrazerebroventrikuläre Infusion (ICV) Darreichungsform, Darreichungsbedingungen und Hinweise/ Safety Monitoring: • entsprechen der Beschreibung der Studie 190-201 (siehe oben)	
Nicht erlaubte Begleitmedikation (gemäß Studienbericht) • Vorherige Enzymersatztherapie (ERT) zur Behandlung der CLN2-Erkrankung • Jegliche Untersuchungsmedikamente innerhalb der ersten 30 Tage nach der ersten Infusion sowie während der gesamten Studie Erlaubte Begleitmedikation (gemäß Studienbericht) • Gabe von Antihistaminen und zusätzlicher Antipyretika (ca. 30 min vor der Infusion von Cerliponase alfa) möglich • Vorbehandlung mit Sedativa vor Infusionsgabe (gemäß Alter der Probanden) • Medikament: Antikonvulsivum • Medikamente gegen: Myoklonus, Tremor, Unruhe, Schmerzen (Prüfer wurden angehalten, die Medikamente über die gesamte Studie konstant weiterzugeben, sofern keine Änderungen aufgrund von Toxizität oder mangelnder Wirksamkeit erforderlich waren.)	

Abkürzungen: CLN2: Ceroid-Lipofuszinose Typ 2; DMC: Datenkontrollkomitee; ERT: Enzymersatztherapie

2.4.2 Studie 190-901

Gemäß Dossier Modul 4 werden die Ergebnisse der Zulassungsstudien 190-201/190-202 mit einer historischen Kontrolle unbehandelter Patienten verglichen. Die Daten der historischen Kontrollgruppe (Studie 190-901) wurden im Rahmen des DEM-CHILD-Patientenregisters erhoben. Das Studiendesign und die Patientenpopulation der Studie 190-901 werden in der Tabelle 3 charakterisiert.

Gemäß dem finalen SAP (Version 3, 25. März 2016) der Studie 190-201/190-202 war für die primäre Auswertung ein Vergleich mit einer historischen Kontrollgruppe vorgesehen. Dieser wurde auf Empfehlung der Zulassungsbehörden im Statistischen Analyseplan der „Integrated Summary of Efficacy“ (SAP ISE) spezifiziert: „Auf Empfehlung der Zulassungsbehörden wurde im ISE SAP eine Responder-Analyse nach 1:1-Paarbildung der Teilnehmer der Studie 190-201 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) als primärer Endpunkt festgelegt“ [2]. Die aktualisierte Analyse erfolgte auf Basis des DEM-CHILD-Datensatzes ‚Supplementary Report der Studie 190-901‘. Das Studiendesign und die Patientenpopulation der Studie 190-901 werden in der Tabelle 3 charakterisiert.

Tabelle 3: Charakterisierung der Studie 190-901

Beschreibung
Original-Datenbank der Studie 190-901 zum natürlichen Verlauf der Erkrankung (Studienbericht: 30. Juni 2015) Zum Zeitpunkt der Einrichtung der Datenbank und den dazugehörigen Kohorten- bzw. registerspezifischen Auswertungen zur Erfassung des natürlichen Verlaufs der CLN2-Erkrankung war die vergleichende Analyse mit den Ergebnissen der Studien 190-201/190-202 nicht geplant. Die Ziele der Datenbank waren: • Bestimmung der Progressionsrate und deren Variabilität unter Einsatz einer krankheitsspezifischen 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 und • Vergleich des Patientenalters und der Krankheitsschwere bei zwei unabhängigen Patientenkohorten. In die Original-Datenbank wurden Patienten aus zwei großen Kohorten in Hamburg (n=30, 29 ausgewertete, Längsschnittdaten) bzw. New York (n=49, Querschnittsdaten; WCMC [Weill Cornell Medical College]) eingeschlossen. Mithilfe der Daten aus Hamburg sollte geklärt werden, wann die Erkrankung klinisch in Erscheinung tritt, wie schnell die Erkrankung voranschreitet und in welchem Alter Symptome auftreten bzw. die Diagnose gestellt wird. Hierfür wurde die krankheitsspezifische Bewertungsskala ‚HML-Skala (Hamburg-Skala)'[1] (Kombinierte Skala aus motorischer Domäne und Sprachdomäne, 0–6 Punkte) eingesetzt. Die WCMC-Kohorte liefert ergänzende quantitative Querschnittsdaten mittels WCMC-Skala[1] (Kombinierte Skala aus den Domänen Gang und Sprache, 0–6 Punkte) der Erkrankung. **Aktualisierte Analyse der Studie 190-901 zum natürlichen Verlauf der Erkrankung – Studie ist Basis für die Kontrollgruppe des historischen Vergleichs (Studienbericht: 01. Mai 2016)** Die aktualisierte Analyse auf Basis eines DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901) dient als historische Kontrolle der Studien 190-201/202, obwohl die vergleichende Analyse mit Einrichtung des DEM-CHILD-Registers nicht geplant war. DEM-CHILD stellt eine Erweiterung des ursprünglichen Datensatzes (Original-Datenbank) aus Hamburg dar. Es handelt sich um eine fortlaufende, multizentrische und multinationale, klinische, europaweite Datenbank mit Sitz in Hamburg, die, verglichen mit der Original-Datenbank, aktualisierte klinische Daten enthält (Datenschnitt für den primären ersten historischen Vergleich: Februar 2015). Die aktualisierte Analyse des DEM-CHILD-Registers (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 der HML-Skala, wie schnell und wie variabel die Erkrankung fortschreitet, nachdem die Patienten symptomatisch geworden sind • Deskriptive Statistik der Patienten-Charakteristika. 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 aus der Original-Datenbank (Studie 190-901). Die aktualisierte Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901) dient als historische Kontrolle der Studien 190-201/202. Bei dieser Analyse wurden spezifische

[1] Im Original bestehen die Hamburg- und die WCMC-Bewertungsskalen aus vier Domänen ‚0–12 Punkte' (Hamburg Skala: motorische Domäne, Sprachkompetenz, Sehvermögen und Krampfanfälle; WCMC: Gang, Sprache, Myoklonus und Ernährung).

Beschreibung
Auswahlkriterien eingesetzt, um eine Patientenpopulation zu erhalten, die den wesentlichen Ein- und Ausschlusskriterien der Studie 190-201 entsprach. Die in der Datenbank identifizierten Patienten mit CLN2 (n=74) wurden mittels verschiedener Filter selektiert, um ein der Studie 190-201 vergleichbares Patientenkollektiv zu erhalten. Von den 74 Patienten aus dem DEM-CHILD-Register hatten 58 Patienten mindestens einen berichteten HML-Score. Darüber hinaus wurden für die 58 verbliebenen Patienten zusätzliche HML-Filter eingesetzt: • Ausschluss von Zwillingen, • Einschluss von Patienten mit mindestens zwei HML-Scores und: o Zwei HML-Scores im Alter von ≥ 36 Monaten, o Ein HML-Score ≥ 3 Punkten, o Zwei HML-Scores zwischen 1 und 5 Punkten und o Mindestens ein Wert ≥ 6 Monate nach der ersten HML-Bewertung. Nach Einsatz der HML-Filter und Ausschluss eines Zwillings verblieben 41 von 74 Patienten (55 %) in der Analyse und erfüllten die Eignungskriterien für den historischen Vergleich mit den Patienten der Studie 190-201/190-202. Von diesen stammten 33 Patienten aus Hamburg (Deutschland) und 8 Patienten aus Verona (Italien). Von den 41 Patienten, die in die erweiterte Analyse der Studie 190-901 eingeschlossen wurden, handelte es sich bei 26 um Patienten der Originalanalyse und bei 15 um neu aufgenommene Patienten (7 in Hamburg und 8 in Verona). Im Modul 4 wird darauf verwiesen, dass es sich bei den Daten, die sich auf den Zeitraum vor Diagnose der CLN2 beziehen, in der Original-Datenbank um retrospektive Angaben handelt und „die meisten Werte nach Diagnosestellung bei den Patienten aus Hamburg tatsächlich gemessen wurden“ [2]. Weiter heißt es: „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). Wie im ISE SAP festgelegt, wurden die imputierten Visiten soweit möglich vor der Analyse aus dem Datensatz entfernt.“ [2]

Abkürzungen: CLN2: Ceroid-Lipofuszinose Typ 2; HML: Hamburg Motor/Language (Kombinierte Skala MotorikSprache, entwickelt in Hamburg); ISE: Integrated Summary of Efficacy; ML: Motor/Language (Kombinierte Skala Motorik/Sprache); MRT: Magnetresonanztomographie; SAP: Statistischer Analyseplan; WCMC: Weill Cornell Medical College

2.5 Endpunkte und statistische Methoden

2.5.1 In der Nutzenbewertung berücksichtige Endpunkte

Tabelle 4: In der Nutzenbewertung berücksichtigte Endpunkte

Endpunkt	Kategorie	Vom pU als patientenrelevant eingestuft	In der Nutzen-bewertung berücksichtigt
Todesfälle[1)]	Mortalität	●	●
Summenscore der adaptierten Hamburger-Skala ‚ML-Skala‘[2) 4)]	Morbidität	●	●
Responderanalyse[3)]		●	●
Slope-Analyse[5)]		●	●
Summenscore der adaptierten MLV-Skala (ML-Skala plus Sehvermögen)		●	○[6)]

Endpunkt	Kategorie	Vom pU als patientenrelevant eingestuft	In der Nutzenbewertung berücksichtigt
Summenscore der adaptierten MLVS-Skala (ML-Skala plus Sehvermögen und Krampfanfälle)		●	○[6)]
PedsQL (Parent Report for Toddlers)[7)]	Lebensqualität	●	●
PedsQL (Family Impact Modul)[8)]		●	-
CLN2-spezifische QoL-Fragebogen		●	-
Unerwünschte Ereignisse (UE)	Sicherheit	●	●
UEs vom CTCAE-Grad ≥ 3		●	●
Schwerwiegende UE (SUE)		●	●
UEs, die zum Studien- oder Behandlungsabbruch führen		●	●
UEs besonderen Interesses[9)10)]		●	●

● Endpunkt wird als bewertungsrelevant eingestuft und in der Nutzenbewertung berücksichtigt.
○ Endpunkt wird nicht als bewertungsrelevant eingestuft, aber ergänzend in der Nutzenbewertung dargestellt.
- Endpunkt wird nicht als bewertungsrelevant eingestuft.

[1)] Das Gesamtüberleben war gemäß Studienprotokoll kein geplanter Effektivitätsparameter. Gleichwohl wurden Todesfälle als Sicherheitsendpunkt (UE) erfasst.
[2)] Die ML-Skala (0–6 Punkte) stellt den primären Endpunkt der Studie dar und umfasst die zwei Domänen ‚motorische Fähigkeiten und Sprachvermögen'.
[3)] Der primäre Endpunkt der Studie 190-201/190-202 war der Anteil der Patienten, die auf die Therapie angesprochen hatten (Responder; ITT). Responder waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 bzw. Woche 80 (Studientag 340 bzw. 540 bezogen auf die erste 300-mg-Infusion) nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme von maximal 1 Punkt).
[4)] Der ursprüngliche SAP der Studie 190-201/190-202 sieht für den primären Endpunkt einen Vergleich gegenüber einer aus den historischen Verlaufsdaten festgesetzten, historischen Kontrollgruppe vor. Dieser wurde erst auf Empfehlung der Zulassungsbehörden im ISE SAP konkretisiert. Der SAP liegt in der finalen 3. Version vom 25. April 2016 vor. In der aktualisierten Analyse der historischen Kontrollgruppe (Supplemental Report der Studie 190-901) wurde die ursprünglich entwickelte Skala ‚HML-Skala' (kombinierte Skala aus motorischer und sprachlicher Domäne) verwendet. Diese wurde für die einarmigen Studien 190-201/190-202 mit den Entwicklern der HML-Skala marginal angepasst, um einerseits objektive Ankerpunkte zu erhalten und andererseits die Abgrenzung zwischen den Kategorien zu präzisieren (siehe 2.5.3).
[5)] Slope-Analyse (Ausmaß des Abfalls (lineare Steigung der Linie) der ML-Skala, absoluter Wert und kategorisiert nach < 2/ ≥ 2 [ursprünglicher primärer Endpunkt]
[6)] Die Symptome der adaptierten MLV- und MLVS-Skala (Sehvermögen und Krampfanfälle) werden als patientenrelevant eingestuft. Allerdings fehlen zu diesen Skalen Angaben zur Validität (siehe Tabelle 6). Vor diesem Hintergrund werden die Domänen Sehvermögen und Krampfanfälle im Rahmen der Erhebungsinstrumente MLV-/MLVS-Skala in der Nutzenbewertung extrahiert, jedoch aufgrund unzureichender Validität nur ergänzend dargestellt.
[7)] Fremdbeurteilung der Lebensqualität mittels PedsQL™-Fragebogen ("Parent Report for Toddlers"). Der ‚Parent report for toddlers' zählt zum Generic-Core-Scales-Paket des PedsQL Inventory. Das Paket beinhaltet die vier Domänen physikalische, emotionale, soziale und schulische Funktion (4 Dimensionen, 21 Items).
[8)] Die Fremdbeurteilung der Lebensqualität mittels PedsQL™-Fragebogenmodul "Family Impact Module" wird gemäß AM-NutzenV als nicht bewertungsrelevant eingestuft, da es sich per se um kein Instrument handelt, das die Lebensqualität vom Patienten adressiert, sondern die Auswirkungen der Lebensqualität auf die Familie widerspiegelt.
[9)] A priori wurden folgende UEs von besonderem Interesse definiert; Überempfindlichkeits- und anaphylaktische Reaktionen, Infusionsassoziierte UEs (TRE), Erkrankungen des Nervensystems, MedDRA-SMQ "Konvulsionen", Status epilepticus, Hydrozephalus und Meningitis, Unerwarteter und rascher, nicht anders erklärbarer Abfall des CLN2-Scores (ML).
[10)] UEs, die mit dem ICV-Zugang in Zusammenhang standen, wurden in der Studie 190-201 als UE von besonderem Interesse berichtet, allerdings wurde dies nicht a priori für die Studie 190-201, sondern erst für die Studie 190-202 definiert (Amendement 3: 26. Februar 2016).

Abkürzungen: CLN2: Ceroid-Lipofuszinose Typ 2; ML: Motor/Language (Kombinierte Skala Motorik/Sprache); PedsQL: Pediatric Quality of Life Inventory; UE: unerwünschte Ereignisse, SUE: schwerwiegende UE; QoL: Quality of life (Lebensqualität)

Abweichend von der Einschätzung des pU wird der CLN2-spezifische QoL-Fragebogen nicht in der vorliegenden Nutzenbewertung berücksichtigt. Es handelt sich hierbei um ein vom pU eigenständig entwickeltes QoL-Instrument, was positiv zu bewerten ist. Allerdings fehlen von Seiten des pU Angaben hinsichtlich der Entwicklung und Validierung, weshalb eine abschließende Bewertung nicht möglich ist. Darüber hinaus wird die Fremdbeurteilung der Lebensqualität mittels PedsQL™-Fragebogenmodul ”Family Impact Module“ gemäß AM-NutzenV als nicht bewertungsrelevant eingestuft, da es sich per se um kein Instrument handelt, das die Lebensqualität vom Patienten adressiert, sondern die Auswirkungen der Lebensqualität auf die Familie widerspiegelt. Darüber hinaus werden die MLV- und MLVS-Skala als nicht bewertungsrelevant eingestuft, jedoch ergänzend in der Nutzenbewertung dargestellt. Es fehlen von Seiten des pU Angaben hinsichtlich der Entwicklung und Validierung für die beiden Skalen, weshalb eine abschließende Bewertung nicht möglich ist (siehe Tabelle 6).

2.5.2 Verzerrungspotential auf Studien- und Endpunktebene

Da es sich bei den Studien 190-201 und 190-202 um Studien ohne Kontrollgruppe handelt, wird von einem hohen Verzerrungspotential auf Studien- und Endpunktebene ausgegangen.

Aufgrund des einarmigen Studiendesigns legt der pU einen historischen Vergleich für den primären Endpunkt – die HML-Skala (Studie 190-901) bzw. die adaptierte ML-Skala (Studie 190-201/190-202) – vor [analysierte Endpunkte: Summenscore der ML-Subskalen, MLV-Skala und MLVS-Skala, siehe 2.5.1]. Für den Vergleich des primären Endpunkts der Studie 190-201/190-202 mit der historischen Kontrollstudie 190-901 wurden prospektiv Matching-Kriterien für die Paarbildung definiert (siehe Kapitel 2.5.4). Aufgrund der per se geringen Ergebnissicherheit eines historischen Vergleichs können nur solche Unterschiede zwischen den gegenüberstehenden Behandlungsergebnissen relevant sein, die eine Größenordnung aufweisen, mit der ausgeschlossen werden kann, dass die beobachteten Unterschiede allein auf systematischer Verzerrung beruhen (weitere Angaben zur Methodik und Validität der Ergebnisse siehe Kapitel 2.5.4 und 2.6.3).

2.5.3 Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz

Laut SGB V § 35a Absatz 1 Satz 10 gilt der medizinische Zusatznutzen durch die Zulassung als belegt. Nachweise gemäß § 5 Absatz 1 bis 6 sind nicht verpflichtend vorzulegen. Zur Beurteilung der Patientenrelevanz der vorgelegten Endpunkte wurden das Dossier, die zugehörigen Studienberichte und die in diesen Quellen zitierte Literatur herangezogen. Ferner wurden eigene Recherchen zur Validität und Patientenrelevanz der Endpunkte durchgeführt. Das Ergebnis der Betrachtung der Endpunkte findet sich aufgeteilt nach Endpunktkategorien in den Tabellen 5 bis 8.

Tabelle 5: Beschreibung der Endpunkte der Kategorie Mortalität

Mortalität	
Endpunkt	**Todesfälle**
Operationalisierung	**Studie 190-201/190-202** Gemäß finalem Studienprotokoll der Studie 190-201 (Amendement 5) und Studienprotokoll der Studie 190-202 (Amendement 3) wurden Todesfälle als schwerwiegendes unerwünschtes Ereignis (SUE) erfasst. Die Erfassung begann mit der Einverständniserklärung und setzte sich über einen Nachbeobachtungszeitraum von 6 Monaten nach der letzten Infusion oder einem vorzeitigen Abbruch fort. Die Nachbeobachtung entfiel, wenn der Patient innerhalb von vier Wochen nach der letzten Infusion in die Folgestudie 190-202 eingeschlossen wurde. Gemäß Modul 4 des Herstellerdossiers konnte das Gesamtüberleben angesichts der Lebenserwartung in den Zulassungsstudien bisher nicht erfasst werden. Nach Einschätzung des pU wird die derzeit laufende Extensionsstudie 190-202 erste Ergebnisse zur Mortalität liefern – definiert als Gesamtrate der Todesfälle zwischen Studienbeginn und Studienende. **Studie 190-901** In der Originalanalyse der Studie 190-901 zum natürlichen Verlauf der Erkrankung (Studienbericht: 30. Juni 2015) wurde das Sterbedatum der Patienten erfasst. Die aktualisierte Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901, 1. Mai 2016) enthält keine Daten zum Gesamtüberleben.
Bewertung	**Operationalisierung:** Die Operationalisierung ist nachvollziehbar beschrieben. Vor dem Hintergrund, dass die Erkrankung zwischen dem 2. und 4. Lebensjahr symptomatisch bzw. diagnostiziert wird [28] und einer prognostizierten Lebenserwartung von 10–15 Jahren [23], sind auf Basis des aktuellsten Datenschnitts eher keine oder kaum Todesfälle zu erwarten. Dennoch bleibt unklar, warum in der aktualisierten Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901, 1. Mai 2016) keine Daten – für einen späteren Vergleich – erfasst wurden. **Validität und Patientenrelevanz:** Der Endpunkt ist valide. Entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (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.“

Tabelle 6: Beschreibung der Endpunkte der Kategorie Morbidität

Morbidität	
Endpunkt	**Summenscore der adaptierten Hamburger-Skala ‚ML-Skala‘ (primärer Endpunkt)**
Operationalisierung	**Studie 190-201/190-202** Der primäre Wirksamkeitsendpunkt der Studie 190-201/190-202 ist der Summenscore der modifizierten motorischen Domäne und Sprachdomäne (modifizierte ML-Skala). Im Studienverlauf wurde die Primäranalyse der Krankheitsprogression, gemessen mittels ML-Skala, von einer Slope- in eine Responder-Analyse geändert (SAP Version 3 vom 25.03.2016). Die Slope-Analyse blieb als Endpunkt bestehen (siehe Kapitel 2.5.4.1).

Morbidität	
	Responder, als Anteil der Patienten, die auf die Therapie angesprochen hatten, waren definiert als Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme von maximal 1 Punkt). Demgegenüber waren Non-Responder (bzw. stabiler Abfall) definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) um zwei oder mehr Punkte (≥ 2) abfiel und bis zum letzten Untersuchungszeitpunkt nicht wieder um einen Punkt (oder mehr) zugenommen hatte. Der ursprüngliche SAP der Studie 190-201/190-202 sieht für den primären Endpunkt einen Vergleich gegenüber Verlaufsdaten einer historischen Kontrollgruppe vor. Dieser Vergleich wurde erst auf Empfehlung der Zulassungsbehörden im ISE SAP konkretisiert, wonach eine Responder-Analyse nach 1:1-Paarbildung (Matching) der Teilnehmer der Studie 190-201/202 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) als primäre Analyse festgelegt wurde. Der SAP liegt in der finalen 3. Version vom 25. April 2016 vor. Zur Erhebung der CLN2-Morbidität mittels der Hamburger Skala wurde in der historischen Kontrollgruppe die ursprünglich entwickelte ‚HML-Skala' (kombinierte Skala aus motorischer und sprachlicher Domäne) verwendet. Für die einarmigen Studien 190-201/190-202 wurde diese mit den Entwicklern der HML-Skala marginal angepasst, um einerseits objektive Ankerpunkte zu erhalten und andererseits die Abgrenzung zwischen den Kategorien zu präzisieren (siehe Abbildung 3). Sowohl die ursprüngliche HML-Skala [24] als auch die geringfügig modifizierte ML-Skala (0–6 Punkte) umfassen zwei Domänen (Motorische Fähigkeiten und Sprachvermögen) von den vier Domänen der MLVS-Gesamtskala (plus Sehvermögen [Vision] und Krampfanfälle [Seizure]). Die motorischen Fähigkeiten und das Sprachvermögen werden auf einer 4-Punkte-Skala bewertet (0–3 Punkte), wobei die einzelnen Abstufungen innerhalb beider Domänen normale Fähigkeiten (3 Punkte) bis hin zum vollständigen Funktionsverlust (0 Punkte) beschreiben. Die modifizierte MLV-Skala (0–9 Punkte) besteht aus den drei Domänen: motorische Fähigkeiten, Sprachvermögen und Sehvermögen, und die modifizierte MLVS-Skala (0–12 Punkte) besteht aus den vier Domänen: motorische Fähigkeiten, Sprachvermögen, Sehvermögen und Krampfanfälle. Diese Domänen werden durch den Prüfarzt erhoben. Ausgewertet wurden diese Skalen gemäß Studienbericht als Veränderung zum 300-mg-Baseline-Wert. Eine Gegenüberstellung der beiden vom pU als primär relevant für die Bewertung der Wirksamkeit eingestuften Domänen (motorische und sprachliche Fähigkeiten) der ursprünglichen HML-Skala und der modifizierten ML-Skala ist nachfolgend in Abbildung 3 dargestellt [2].

Morbidität

HML-Skala [5]			ML-Skala [37]
Motorische Fähigkeiten (Hamburg)	3	Geht normal[1]	Weitgehend normaler Gang. Keine auffällige Ataxie, keine krankheitsbedingten Stürze.
	2	Häufige Stürze, offensichtliche Ungeschicklichkeit	Selbständiges Gehen, gekennzeichnet durch die Fähigkeit, ≥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
Sprachver-mögen(Hamburg)	3	Normal (individuell beste Leistung)[2]	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)[3]
	1	Schwer verständlich	Schwer verständlich. Nur wenige verständliche Worte.
	0	Unverständlich oder keine Sprache	Keine verständlichen Worte oder Vokalisation

[1] Bei einigen Kindern war die motorische Entwicklung nie altersadäquat.

[2] 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 vergeben.

[3] 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.

Abbildung 3: Unterschiede zwischen den Domänen ‚motorische und sprachliche Fähigkeiten' zwischen der ursprünglichen HML-Skala (verwendet in historischer Kontrollstudie 190-901) und der modifizierten ML-Skala (verwendet in der Studie 190-201/190-202 [2]).

Standardisierung der Bewertungen

Die Bewertung der Skala erfolgt durch sogenannte „Rater" (nachfolgend als Prüfer bezeichnet), die vor Studienbeginn ein Training absolvieren müssen, um eine möglichst standardisierte Erhebung und Bewertung der Ergebnisse in Übereinstimmung mit dem Manual während der Studie zu gewährleisten. Das Training beinhaltet 3–5 Trainingsfälle, die vom Prüfer bewertet werden und durch den Trainer als akzeptabel eingestuft werden mussten. Nach Möglichkeit sollten Patienten über den gesamten Studienverlauf von ein und demselben Prüfer bewertet werden. Die Erhebungen wurden in den Studienzentren regelmäßig überwacht um sicherzustellen, dass die Tests im Verlauf der Studie einheitlich erhoben werden. Die Bewertungen sollten dabei vor der Applikation der Infusion bzw. Studienuntersuchung (MRT-Untersuchung sowie der Blut- und CSF-Probenahme) stattfinden [10].

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 erfolgen. Falls der Prüfarzt die Muttersprache des Patienten nicht beherrscht, stand ein zertifizierter Übersetzer bereit, der bewertete, ob die Sprache des Patienten verständlich war. Auch für die Eltern- /Betreuerinterviews konnte ein Übersetzer herangezogen werden [10].

Alle Tests wurden mittels Videokamera aufgezeichnet (geändert mit Amendement 3 des Studienprotokolls, 07.07.2014, ursprünglich alle 24 Wochen). Die Videoaufnahmen

Morbidität	
	waren in Übereinstimmung mit dem Manual “Ratings Assessment Guidelines” standardisiert. Die Videoaufzeichnungen wurden nicht zur Bewertung herangezogen, sondern wurden ausschließlich für die klinischen Beobachtungen genutzt [10]. Zusätzlich zu den klinischen Bewertungen durch die Prüfärzte wurden die Eltern/Betreuer befragt. Diese wurden gebeten, Angaben zur Gehfähigkeit, Sprache, zum Sehvermögen, zu anomalen unwillkürlichen Bewegungen und zur Ernährung der Kinder in der vorangegangenen Woche zu machen. Im Verlauf der Interviews stellten die Prüfärzte zu jeder Fähigkeit gezielte Fragen. Die Interviews wurden nicht zur Erhebung des ML-Scores bzw. MLVS-Scores herangezogen, sondern dienten dazu, eventuelle Unstimmigkeiten verglichen mit den Beobachtungen des Prüfarztes aufzudecken oder wie der Patient am besten motiviert werden kann [10]. <u>Praktische Ausführung der Erhebung [10,31]</u> *Domäne motorische Fähigkeiten* • Mündliche Aufforderung des Patienten, sich im Rahmen seiner Möglichkeiten mit oder ohne Hilfsmittel fortzubewegen z. B. durch den Raum zu gehen oder vom Tisch zum Stuhl und zurück (etc.) • Hilfsmittel umfassten u. a. Wände, Möbel oder Eltern/Betreuer, die zur Fortbewegung eingesetzt/genutzt wurden. • Beobachtet wurden alle spontanen Bewegungen, das Gangbild/Ataxien, Stürze, die Fähigkeit, aus dem Sitzen aufzustehen sowie der Einsatz externer Hilfsmittel. • Darüber hinaus wurden 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? *Domäne Sprachvermögen* • Der Prüfarzt bewertete das Sprachvermögen des Patienten einschließlich verständlicher Wörter und grammatikalisch korrekter Sätze. Mitberücksichtigt wurde dabei die Qualität der Äußerungen. • Erfasst wurde u. a., ob der Patient 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. Genutzt wurden hierfür bevorzugt dem Patienten vertraute Gegenstände, um das Sprachvermögen zu überprüfen. • Darüber hinaus wurden Eltern/Betreuer 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? *Sehvermögen (ML-Skala)* • 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. • 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.

Morbidität	
	Krampfanfälle (ML-Skala) • 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 (> 1 pro Woche). Erhebungszeitpunkte [10] Die zwei Domänen der Skala wurden zm Screening, zu Baseline vor der ersten Infusion und an Tag 4 nach der ersten Infusion erhoben. Während der Dosis-Eskalationsphase wurden die Domänen alle 4 Wochen und an Tag 4 nach jeder Dosis-Eskalation erhoben. Im Rahmen der stabilen Dosisphase wurde die Skala für alle Patienten in Woche 1 sowie alle 8 Wochen und bei vorzeitigem Studienabschluss oder vorzeitigem Studienabbruch erhoben. Baseline-Wert [12] Letzte Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet). **Studie 190-901 (Studie zum natürlichen Krankheitsverlauf [historische Kontrollgruppe])** Wie bereits oben beschrieben, wurde in der Studie 190-901 die Krankheitsprogression mittels HML-Skala erhoben. Gemäß Modul 4 wurde der HML-Score nicht standardisiert erhoben, da keine Erhebungszeitpunkte bei den Kohorten zum natürlichen Krankheitsverlauf bzw. im DEM-CHILD festgelegt wurden. Nach Angabe des pU im Modul 4 wurden die Daten teilweise retrospektiv bzw. die aktuelleren Daten prospektiv erhoben. Datenschnitt war Februar 2015 (Zeitpunkt des Datentransfers von DEM-CHILD).
Bewertung	**ML-Skala (190-201/190-202) und HML-Skala (190-901)** Operationalisierung: • Bei der ML-Skala handelt es sich um eine modifizierte Form der Original-Hamburg-ML-Skala (HML-Skala), bestehend aus den zwei Domänen “motorische Fähigkeiten und Sprachvermögen” der Hamburg-Gesamtskala (HMLVS mit insgesamt vier Domänen: motorische Fähigkeiten, Sprachvermögen, Sehvermögen und Krampfanfälle). • Die Begründung für das Ausklammern der Domänen Sehvermögen und Krampfanfälle ist nicht nachvollziehbar. Der pU begründet dies damit, dass sowohl die Krampfanfälle als auch das Sehvermögen in der Studie 190-201/190-202 keine Berücksichtigung fanden, da diese im Gegensatz zu den anderen zwei Domänen als weniger sensitiv eingestuft werden. Der pU argumentiert, dass Krampfanfälle über den Krankheitsverlauf hinsichtlich ihrer Häufigkeit und Schwere variieren und darüber hinaus mit Hilfe von Begleitmedikationen moduliert werden. Die Verschlechterung des Sehvermögens tritt erst im späteren Krankheitsverlauf auf und scheint laut pU bis zum 10. Lebensjahr nicht offensichtlich und diagnostizierbar. Der pU führt aus, dass die zwei Domänen keine ausreichenden sensitiven Indikatoren für die Beurteilung der Krankheitsprogression darstellen [10]. Dieser Einschätzung kann nicht gefolgt werden (siehe unten Patientenrelevanz).

Morbidität	
	Validität: Übereinstimmung zwischen der ursprünglichen Hamburg-Skala (motorische und sprachliche Funktion [HML-Skala]) und der modifizierten ML-Skala: Im Rahmen der vom pU eingereichten sogenannten Brückenstudie [31] „A Brief Clinician-Reported Outcome Evidence Dossier Prepared For: The 0-6 Hamburg Scale in Patients with Neuronal Ceroid Lipofuscinosis Type 2 (CLN2)" wurden die beiden Domänen der ursprünglichen Hamburg-Skala (HML-Skala) für die Studie 190-201 verändert, um einerseits objektive Ankerpunkte zu erhalten und andererseits die Abgrenzung zwischen den Kategorien zu präzisieren. Hierfür wurden anhand von Videoaufnahmen die Ratings zu beiden Domänen der ursprünglichen Hamburg-Skala (motorische und sprachliche Funktion [HML-Skala]) überprüft und ggf. angepasst. Die vorgenommenen Modifikationen waren gering und betrafen primär die motorische Domäne (z. B. eine Punktzahl von 2 ist definiert als „häufige Stürze, offensichtliche Ungeschicklichkeit" und wurde modifiziert in „abnorme Gangart, unabhängig ≥ 10 Stufen, häufige Stürze, offensichtliche Ungeschicklichkeit"). Zur Überprüfung der Übereinstimmung zwischen den Bewertungen durch den unabhängigen Gutachter unter Einsatz der HML-Skala und denjenigen der Prüfärzte der Studie 190-201/202 auf Basis der ML-Skala wurde Cohens Kappa („weighted kappa", Kω) als statistisches Maß eingesetzt. Im Auftrag vom pU wurde zudem die Validität der ML-Skala sowie deren Einzeldomänen als auch gegenüber der HML-Skala verglichen. Im Ergebnis zeigten sich folgende Übereinstimmungen zu Baseline, Woche 25, 48 und 72: zwischen den beiden motorischen Skalen (0,67 ≤ Kω ≤ 1,00) und zwischen den beiden Sprachskalen (0,34 ≤ Kω ≤ 0,62) und bezogen auf die Summenskala aus beiden Domänen (0,67 ≤ Kω ≤ 0,79). Darüber hinaus ergaben die Analysen der mittleren Steigung („slope"), berechnet über den Zeitraum zwischen Baseline und Woche 48, nahezu ähnliche Punktschätzer für die HML- und ML-Skala (0,14±0,25 [HML] vs. 0,17±0,33 [ML]), die Sprachskala (0,00±0,00 [HML] vs. -0,02±0,575 [ML]) und die Summenskala aus beiden Domänen (0,05±0,41 [HML] vs. 0,17±0,52 [ML]). Zur Reliabilität und Validität der Hamburger Gesamtskala mit 4 Domänen (HMLVS) finden sich in den vom pU eingereichten Daten sowie im Rahmen einer eigenen orientierenden Recherche keine Validierungsstudien, die Angaben zum Entwicklungsprozess und zur Validierung machen. Folglich ist eine abschließende Bewertung der (H)ML- und Gesamt-Skala auf dieser Basis nicht möglich. Zwei andere Studien von Worgall und Kollegen aus dem Jahr 2007 [29] und 2008 [30] nutzten eine entsprechende HMLS-Skala bestehend aus den drei Domänen motorische und sprachliche Fähigkeiten und Krampfanfälle. Auch hier fehlen Angaben zum Entwicklungsprozess und zur Validierung. Hinsichtlich der Reliabilität wurde vom pU zum einen die interne Reliabilität anhand der Baselinedaten aus der pivotalen Studie 190-201 mittels des Spearman-Brown-Korrelationsmaßes abgeschätzt (r=0,79). Trotz fehlender Evaluation in separater Studie wird dieser Wert als ausreichend bewertet. Zum anderen wurden in einer formalen Inter-Rater-Studie insgesamt 36 Videoaufzeichnungen von 12 der in Hamburg rekrutierten Patienten der Studie 190-201/190-202 zu vier Zeitpunkten (Baseline, Woche 25, 48 und 72 [Woche 25 der Studie 190-202]) von einem Rater nochmals beurteilt. Die Übereinstimmung wurde mittels gewichtetem Kappa analysiert. Im Ergebnis zeigte sich für die motorischen Scores eine Übereinstimmung von Kω=0,93 für alle 36 Videos (jeweils Kω=1,00 in Woche 25 (n=11), Woche 48 (n=10) und Woche 72 (n=3) bzw. Kω=0,76 für Baseline (n=12), da zwei der motorischen Skalenwerte zu Baseline durch den Rater einen Punkt niedriger als durch die Prüfärzte bewertet wurden). Für die Scores der Sprachdomäne lag die Übereinstimmung bei Kω=0,82 für alle Videos. Die Kω-Schätzungen schwankten über alle Erhebungszeitpunkte zwischen

Morbidität	
	0,67 in Woche 48 und 0,93 zu Baseline. Für den kombinierten ML-Score stimmten die Ergebnisse zwischen dem Rater und den Prüfärzten zu allen Untersuchungszeitpunkten gut überein (0,89 ≤ Kω ≤ 0,93). Dies kann als gute Inter-Rater-Reliabilität hinsichtlich der ML-Scores bewertet werden [31]. Hinweise zur Abschätzung der Inhaltsvalidität können Korrelationsdaten der Studie 190-201/190-202 liefern, die der pU vorlegt. Der kombinierte ML-Score bei Baseline und deren zwei Einzelskalen wurde mit folgenden PRO-Endpunkten bei Baseline verglichen: PedsQL™-Generic-Core Gesamt-Score des Moduls „Parent Report for Toddlers"; PedsQL™-Gesamt-Score des Moduls „Family Impact Modul [Familiäre Belastung]" und dem vom pU entwickelten CLN2-spezifischen QoL-Fragebogen [31]. Die beiden letzteren Endpunkte werden in dieser Nutzenbewertung jedoch nicht als bewertungsrelevant eingestuft. Vom pU wurden für die motorische Domäne der ML-Skala nur diejenigen Korrelationen angegeben, die der pU als maßgeblich für einen Zusammenhang mit der Motorik ansah: Gesamt- und Physische Unterskala des PedsQL™ Toddler-Kernmoduls, der Unterskalen „Körperliche" bzw. „Soziale" Funktion des Moduls „Familiäre Belastung" sowie der Unterskala „Tägliche Aktivitäten" des CLN2-QoL-Fragebogens. Diese Korrelationen waren moderat (0,40 ≤ r ≤ 0,64), wobei die Korrelation mit dem Endpunkt „Körperliche Funktion" (PedsQL™ Generic-Core für Toddlers) am höchsten war. Auch die Korrelationen zwischen den Baseline-Werten zum Sprachvermögen und den Unterskalen „Soziale", „Kognitive" Funktion und „Kommunikation" des Moduls „Familiäre Belastung" des PedsQoL-Moduls waren moderat (0,47 ≤ r ≤ 0,64; vom pU ausgewählte und berichtete Werte). Zwischen dem Gesamtscore der modifizierten ML-Skala (0–6) und dem „Toddler" PedsQL™-Generic-Core Gesamt-Score, dem „Familiäre Belastung"-PedsQL™-Gesamt-Score sowie der Unterskala „Tägliche Aktivitäten" des CLN2-QoL-Fragebogens ergaben sich geringe bis moderate Korrelationen (0,35 ≤ r ≤ 0,51; vom pU ausgewählte und berichtete Werte) [31]. Zur Beurteilung der Änderungssensitivität legt der pU Korrelationsanalysen aus der pivotalen Studie 190-201 vor. Die ML-Skala sowie deren Einzel-Scores wurden anhand des Korrelationskoeffizienten nach Spearman untersucht. Die Änderung der ML-Scores zwischen Baseline und Woche 48 (Studienende/vorzeitiger Abbruch) wurden mit den Änderungen der Domäne „Alltagsaktivitäten" des CLN2-QoL-Fragebogens 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 Korrelation (r=0,37 bzw. 0,32, n=22). Die Korrelation zwischen der Änderung des PedsQL™ Domäne „Soziale Funktion" und dem Sprachvermögen, gemessen auf der modifizierten ML-Skala, betrug 0,29. <u>Abschließende Bewertung der Validität und Patientenrelevanz</u> Die oben angeführten Daten und Angaben entsprechen nicht einer separaten und unabhängigen Validitätsuntersuchung, d. h. einer Überprüfung außerhalb der (pivotalen) Studie zur Evaluation von Therapieeffekten. Eine methodisch adäquate Validierung des Instruments liegt nur eingeschränkt vor. Deshalb ist eine eindeutige und abschließende Bewertung nicht möglich. Die ML-Skala kann jedoch als etabliertes Instrument in der Patientenpopulation gesehen werden, was auch Experten bestätigen. Die Reliabilität scheint aufgrund der Angaben gegeben. Die Hinweise zur Abschätzung einer Inhaltsvalidität erscheinen für die modifizierte und originale ML-Skala plausibel und unter Berücksichtigung der beschriebenen methodischen Einschränkungen akzeptabel, wobei hinsichtlich der Änderungssensitivität die meisten Bedenken bestehen. Unter Berücksichtigung des natürlichen Krankheitsverlaufs bei Patienten mit bestätigter CLN2-Erkrankung erscheint ein Erhalt oder eine Verbesserung dieser in der ML-Skala adressierten motorischen und

Morbidität	
	sprachlichen Fähigkeiten als nachvollziehbar und patientenrelevant. Der primäre Endpunkt wird nachfolgend dargestellt und für die Bewertung des Zusatznutzens von Cerliponase alfa berücksichtigt. Für die anderen beiden Domänen (Krampfanfälle und Sehvermögen) bzw. für die (H)MLV- und (H)MLVS-Gesamtskalen fehlen Angaben zum Entwicklungs- und Validierungsprozess, so dass hinsichtlich der Validität keine Einschätzung möglich ist. Da jedoch das Sehvermögen und Krampfanfälle wichtige und bekannte Aspekte des natürlichen Verlaufs der CLN2 darstellen, werden die Domänen Sehvermögen und Krampfanfälle als patientenrelevant eingestuft. In der Folge werden diese in der vorliegenden Nutzenbewertung ausschließlich ergänzend herangezogen und als nicht bewertungsrelevant berücksichtigt. Entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (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.“

Tabelle 7: Beschreibung der Endpunkte der Kategorie Lebensqualität

Morbidität	
Endpunkt	**PedsQL™ Module: „Parent report for toddlers“**
Operationali-sierung	**Studie 190-201** Bei dem PedsQL™-Modul:-„Parent report for toddlers“-Fragebogen handelt es sich um ein generisches Instrument, welches von Angehörigen beantwortet wird (Proxy). Die Erhebung der Lebensqualität wurde nicht während der Dosis-Eskalationsphase, sondern ausschließlich während der stabilen Dosisphase erhoben. Als Baseline-Wert für die Bewertung der Lebensqualität mittels PedsQL™ (Module: „Parent report for toddlers“ und „Parent family impact“) wurde grundsätzlich die letzte Beobachtung vor der ersten 300-mg-Infusion bestimmt (max. ≤ 8 Wochen vorher); lag diese nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion eingegeben. Erhebungszeitpunkte Daten zur Lebensqualität wurden zu Baseline (vor der ersten Infusionsgabe der stabilen Dosisphase; Definition siehe oben) und danach alle 12 Wochen sowie zu Studienende oder bei vorzeitigem Abbruch erhoben. Die Fragebögen sollten möglichst vor den MRIs und den Probenahmen ausgefüllt werden. Eine Imputation für fehlende Werte war nicht vorgesehen. Der “Parent Report for Toddlers” ist ein Fremdbeurteilungsinstrument zur Erfassung der Lebensqualität bei pädiatrischen Populationen mit chronischen Erkrankungen bei Kindern zwischen 2 und 4 Jahren. Dieser setzt sich aus 21 Items, die vier Dimensionen abdecken, zusammen: • Körperliche Funktionen (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 Funktionen (5 Items: Angstgefühle, Traurigkeit oder Depressionen, Schlafprobleme und Sorgen) • Soziale Funktionen (5 Items: Probleme, mit anderen Kindern zu spielen, andere Kinder wollen nicht mit ihr oder ihm spielen, von anderen Kindern geärgert werden, nicht in der Lage sein, Dinge zu tun, die andere Kinder ihres/seines Alters können, beim Spiel mit anderen Kindern mithalten)

Morbidität	
	• Schulische Funktionen (3 Items: Probleme, in der Schule mit den Schulkameraden mitzuhalten, Fehltage in der Schule oder dem Kindergarten aufgrund von Unwohlsein oder Arztbesuchen/Krankenhausauenthalten). 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. Scoring nach Dimension: Der Skalen-Score wurde nicht berechnet, wenn mehr als 50 % der Items der Skala fehlten. 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 Fähigkeiten sowie schulische Fähigkeiten. Gesamtscore: Summe der beantworteten Items aus allen Dimensionen. Gemäß SAP ist vorgesehen, die unterschiedlichen Scores nach Dimension, Psychosozialer Gesundheits-Summenscore und Gesamtscore deskriptiv zusammenzufassen.
Bewertung	**Operationalisierung:** Der ‚Parent report for toddlers' zählt zum Generic-Core-Scales-Paket des PedsQL-Inventory. Die allgemeine Kernskala ‚Parent report for toddlers' (23 Items) beinhaltet die 4 Domänen körperliche (8 Items), emotionale (5 Items), soziale (5 Items) und schulische Funktion (5 Items). Neben den vier Einzelskalen können drei unterschiedliche Summenskalen gebildet und ausgewertet werden, der ‚Psychosocial Health Summary Score', der Gesamtscore sowie der ‚Physical Health Summary Score', der jedoch der Domäne ‚körperliche Funktion' entspricht (Varni et al. 1999). Es liegen Versionen zur Selbstbeurteilung für Vorschulkinder (5 bis 7 Jahre), Schulkinder (8 bis 12 Jahre) und Jugendliche (13 bis 18 Jahre) sowie Elternversionen für diese Altersgruppen und Kleinkinder (2 bis 4 Jahre) vor [21]. In der vorliegenden Studie 190-201/ 190-202 wurde die Elternversion für Kinder zwischen 2 und 4 Jahren ("Parent Report for Toddlers") verwendet. Hierbei handelt es sich um ein Fremdbeurteilungsinstrument zur Erfassung der Lebensqualität bei pädiatrischen Populationen mit chronischen Erkrankungen. Dieser setzt sich wie eingangs beschrieben aus vier Dimensionen zusammen, beinhaltet allerdings nur 21 anstatt der beschriebenen 23 Items (nur 3 statt 5 Items in Domäne Schulische Funktion). **Validität:** Der PedsQL™ ist ein etabliertes Instrument und wurde in den 1990er Jahren als generisches Instrument zur Erfassung der Lebensqualität bei pädiatrischen Populationen mit chronischen Erkrankungen entwickelt [27]. Seither wurden zahlreiche erkrankungsspezifische Module entwickelt. Der PedsQL™ ist ein modulares Konzept zur Erfassung der gesundheitsbezogenen Lebensqualität von gesunden Kindern und Jugendlichen und jenen mit akuten und chronischen Erkrankungen [27]. Das aus den USA stammende Messinstrument bietet die Möglichkeit des Elternreports für 2- bis 18-Jährige und die des Selbstreports für Kinder und Jugendliche von 5 bis 18 Jahren. In der Studie 190-201/190-202 wurde ausschließlich die Elternversion für Kinder zwischen 2 und 4 Jahren verwendet. Für die Studie 190-201/202 galt das Einschlusskriterium Alter zwischen 3 und 15 Jahren. Das mediane Alter bei Studieneinschluss lag bei 4 Jahren, die Spanne zwischen 3–8 Jahren. Es bleibt unklar, weshalb der pU für Kinder anderer Altersklassen in der Studie nicht andere bestehende Versionen, wie bspw. die Selbstbeurteilungsversion für Vorschulkinder (5 bis 7 Jahre), eingesetzt hat oder weshalb der pU für alle eingeschlossenen Kinder, unabhängig vom Alter, die

Morbidität	
	Elternversion für Kinder zwischen 2 und 4 Jahren verwendete. Die Validität der Ergebnisse ist in dieser Hinsicht als unsicher anzusehen. Für diese Fragebögen konnten für die Evaluation der psychometrischen Eigenschaften gute Ergebnisse berichtet werden [26], so dass das Instrument als valide angesehen werden kann. Der PedsQL™ wurde in diverse Sprachen übersetzt. Die interne Konsistenz des PedsQL™ wurde an einer Stichprobe mit 8.836 gesunden und 847 chronisch kranken Kindern untersucht (Cronbach`s Alpha 0,89 Kinderfragebogen; Cronbachs‘s Alpha 0,92 Elternfragebogen) [25]. Die Test-Retest-Reliabilität bei Kindern (7–12 Jahre) wird mit 0,86 angegeben [16]. Die interne Validität wurde an einer Stichprobe mit ca. 5.500 Kindern (ca. 570 davon chronisch krank) untersucht. Es konnte gezeigt werden, dass Kinder mit chronischen Erkrankungen konsistent eine niedrigere Lebensqualität berichten als gesunde Kinder. Zur Änderungssensitivität liegen weitere Untersuchungen vor. Die MID wird mit 4,4 (Änderung im Gesamtscore für selbstberichtete Ergebnisse) bzw. mit 4,5 (Elternfragebogen) angegeben [25]. **Patientenrelevanz:** Die Erfassung der allgemeinen Lebensqualität anhand des ausreichend validierten Instruments PedsQL 4.0 Generic Core Scales wird als patientenrelevant bewertet (vorausgesetzt die Auswahl der adäquaten Versionen). Der PedsQL Generic Core ist jedoch kein krankheitsspezifisches Instrument und für die CLN2 liegt kein krankheitsspezifisches Modul vor. Entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (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.“

Tabelle 8: Beschreibung der Endpunkte der Kategorie Sicherheit

Sicherheit	
Endpunkt	**Unerwünschte Ereignisse (UE)**
Operationali-sierung	**Studie 190-201** Alle Sicherheitsanalysen wurden mit Daten der Sicherheitspopulation durchgeführt. Gemäß Studienbericht beschränken sich UE auf „treatment emergent adverse events“ (TEAE). Demnach ist ein TEAE definiert als ein neu aufgetretenes oder häufiger auftretendes UE oder sich verschlechterndes UE nach Implantation des ICV-Ports, unabhängig davon, ob der Patient Cerliponase alfa erhalten hatte oder nicht. Vorzugsweise sollte eine Diagnose statt einzelner, mit dieser Diagnose in Beziehung stehender UEs berichtet werden. Die Klassifizierung erfolgte nach Systemorganklasse und Preferred Terms gemäß MedDRA (Version 18.1). UEs (NCI CTCAE) Der Prüfarzt bewertet den Schweregrad jedes Ereignisses nach NCI CTCAE, Version 4. UEs ohne CTCAE-Term wurden entsprechend der allgemeinen CTCAE-Kriterien (Version 4) klassifiziert: • Grad 1: Mild (asymptomatische oder leichte Symptome; klinische oder diagnostische Beobachtungen; Intervention nicht notwendig) • Grad 2: Moderat (minimale, lokale oder nicht-invasive Intervention notwendig; Einschränkung in den instrumentellen Aktivitäten des täglichen Lebens (ADL))

Sicherheit	
	• Grad 3: Schwerwiegend oder medizinisch significant, aber nicht unmittelbar lebensbedrohlich (Krankenhausaufenthalt oder Verlängerung des Krankenhausaufenthalts; Behinderung; Einschränkung Selbstversorgung (ADL)) • Grad 4: Lebensbedrohliche Konsequenzen, dringende Intervention erforderlich, • Grad 5: Tod durch UE Falls ein UE mehr als einmal auftrat, wurde das jeweils schwerste Ereignis registriert. Gemäß SAP wurden darüber hinaus folgende Gruppen von UEs erfasst: • UEs, die gemäß Prüfarzt im Zusammenhang mit der Studienmedikation und/oder ICV-Port stehen (Hinweis: wurde in der Studie 190-201 als UE von besonderem Interesse berichtet, allerdings wurde dies nicht a priori für die Studie 190-201 definiert, sondern erst für die Studie 190-202) • UEs, die zum Tod führen • UEs, die sich nicht zurückgebildet haben oder bei denen Folgeschäden zurückblieben • UEs, die zu einer Studienunterbrechung und/oder vorzeitigem Studienabbruch führen • UEs, die zu einer Dosisreduzierung führen Erhebungszeitpunkte UEs wurden ausschließlich nach der Implantation eines ICV-Ports erfasst. Die Erfassung von UEs beginnt mit der Einverständniserklärung und wird über einen Zeitraum von sechs Monaten nach der letzten Verabreichung des Arzneimittels fortgeführt oder bei vorzeitigem Studienabbruch. Falls der Zeitpunkt des Auftretens eines UEs unbekannt war, wurde das UE als behandlungsbezogen eingestuft. Gemäß Modul 4 im Dossier entfiel die Nachbeobachtung, wenn der Patient innerhalb von vier Wochen nach der letzten Infusion in die Folgestudie 190-202 eingeschlossen wurde. **Studie 190-202** • Die Sicherheit und Verträglichkeit wurde wie oben beschrieben erhoben, abgesehen von folgenden Änderungen: • UEs wurden ausschließlich nach der ersten Dosierung erfasst und über einen Zeitraum von sechs Monaten nach der letzten Verabreichung des Arzneimittels fortgeführt oder bei vorzeitigem Studienabbruch • Alle UEs, die mit dem ICV-Zugang in Zusammenhang standen, wurden als UEs besonderen Interesses definiert (Amendement 3: 26. Februar 2016). Die übrigen UEs besonderen Interesses entsprachen denjenigen der Studie 190-201. **Studie 190-901** • Der Endpunkt wurde in der Studie 190-901 nicht erhoben.
Bewertung	**Operationalisierung** Die Relevanz und Operationalisierung ist nachvollziehbar beschrieben. **Validität und 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.“

Endpunkt	Schwerwiegende unerwünschte Ereignisse (SUE)
Operationali-sierung	**Studie 190-201** Alle Sicherheitsanalysen wurden mit Daten der Sicherheitspopulation durchgeführt. Ein UE galt als schwerwiegend (SUE), sofern es: • als fatal oder lebensbedrohlich einzustufen war, • eine Hospitalisierung oder eine bereits bestehende und verlängerte Hospitalisierung nach sich zog, • zu anhaltender oder bedeutender Einschränkung oder Behinderung führt, • es sich um eine angeborene Anomalie oder einen Geburtsfehler eines Kindes oder Fötus nach Exposition mit dem Prüfpräparat vor der Konzeption oder während der Schwangerschaft oder um ein medizinisch bedeutsames Ereignis oder eine Reaktion handelt. Ein lebensbedrohliches Ereignis bedeutet in diesem Zusammenhang, dass der Patient aufgrund des Ereignisses zu versterben drohte, jedoch nicht die Ereignisse, die hypothetisch zum Tode hätten führen können, wären sie schwerer gewesen. SUEs (NCI-CTCAE-Grad 4 oder 5) UEs mit einem Schweregrad nach NCI-CTCAE-Grad 4 oder 5 wurden als SUE dokumentiert. SUEs vor der Implantation eines ICV-Ports Nach der Einverständniserklärung, aber vor der Implantation eines ICV-Ports wurden ausschließlich SUEs dokumentiert, die im Zusammenhang mit der protokollgesetzten Intervention stehen. Nicht berücksichtigte SUEs umfassten geplante Hospitalisierung für ein Studienverfahren oder elektive Operation vor Aufnahme in die Studie. Nach der Implantation wurden alle SUEs dokumentiert. SUEs (inkl. ICV-Port-SUEs) Alle SUEs (inkl. ICV-Port-SUEs), die entweder im oder nicht im Zusammenhang mit der Studienmedikation standen, waren der Pharmakovigilanz innerhalb von 24 Stunden zu berichten. Dies betraf alle zur Infusion benötigten Komponenten. Die Prüfärzte sollten jeden 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. Gemäß SAP wurden analog zu den UEs ebenfalls folgende Gruppen von SUEs erfasst: • SUEs, die gemäß Prüfarzt im Zusammenhang mit der Studienmedikation und/oder ICV-Port stehen • SUEs, die zum Tod führen • SUEs, die sich nicht zurückgebildet haben oder bei denen Folgeschäden zurückblieben • SUEs, die zu einer Studienunterbrechung und/oder vorzeitigem Studienabbruch führen • SUEs, die zu einer Dosisreduzierung führen Erhebungszeitraum Der Berichtszeitraum für SUEs war wie folgt: • Nach der Einverständniserklärung nur (S)UEs, die mit studienveranlassten Maßnahmen in Zusammenhang standen • ab Implantation des ICV-Zugangs: alle (S)UEs Die Erfassung von SUEs beginnt mit der Einverständniserklärung und wird über einen Zeitraum von sechs Monaten nach der letzten Verabreichung des Arzneimittels fortgeführt oder bei vorzeitigem Studienabbruch. Gemäß Modul 4 im Dossier entfiel die

	Nachbeobachtung, wenn der Patient innerhalb von vier Wochen nach der letzten Infusion in die Folgestudie 190-202 eingeschlossen wurde. **Studie 190-202** Die Sicherheit und Verträglichkeit wurde wie oben beschrieben erhoben, abgesehen von folgenden Änderungen: • SUEs wurden ausschließlich nach der ersten Dosierung in der Studie 190-202 erfasst und über einen Zeitraum von sechs Monaten nach der letzten Verabreichung des Arzneimittels fortgeführt oder bei vorzeitigem Studienabbruch. **Studie 190-901** Der Endpunkt wurde in der Studie 190-901 nicht erhoben.
Bewertung	**Operationalisierung** Die Relevanz und Operationalisierung ist nachvollziehbar beschrieben. **Validität und 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.“
Endpunkt	**UE von besonderem Interesse**
Operationali-sierung	**Studie 190-201/190-202** Operationalisierung: A priori definierte UEs besonderen Interesses („adverse events of special interest“): • Überempfindlichkeitsreaktionen: Standardisierte MedDRA-Abfragen („Standardised MedDRA query“, SMQ) Überempfindlichkeit und anaphylaktische Reaktionen. o 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. • Reaktionen, die in zeitlichem Zusammenhang mit der Infusion auftraten („infusion-associated reaction“, IAR): 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 • Status epilepticus (MedDRA PTs: Petit mal Epilepsie, Status epilepticus) • Hydrozephalus (kommunizierender und nicht kommunizierender): MedDRA PT Hydrozephalus, kongenitaler Hydrozephalus • Meningitis: Entsprechende MedDRA PTs waren im SAP Version 3 definiert. • Unerwarteter, rascher, nicht anders erklärbarer Abfall des CLN2-Scores: Identifiziert durch klinische Bewertung • UEs, die gemäß Prüfarzt im Zusammenhang mit dem ICV-Port stehen (*Hinweis: wurde in der Studie 190-201 als UE von besonderem Interesse berichtet, allerdings wurde dies nicht a priori für die Studie 190-201 definiert, sondern erst für die Studie 190-202)* **Studie 190-901** • Der Endpunkt wurde in der Studie 190-901 nicht erhoben.

Bewertung	Operationalisierung Die Relevanz und Operationalisierung ist nachvollziehbar beschrieben. Validität und 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.“

Abkürzungen: CLN2: Ceroid-Lipofuszinose Typ 2; ML: Motor/Language (Kombinierte Skala Motorik/Sprache); PedsQL: Pediatric Quality of Life Inventory; UE: unerwünschte Ereignisse, SUE: schwerwiegende UE

2.5.4 Statistische Methoden

In diesem Abschnitt werden die geplanten statistischen Methoden zur Auswertung der einarmigen Studien 190-201/190-202, der historischen Kontrollstudie 190-901 und der Vergleich der Wirksamkeitsdaten (ML-/HML-Skala) von Studie 190-201/190-202 vs. Studie 190-901 (Integrated Summary of Efficacy, ISE) beschrieben.

Der finale SAP (Version 3.0 vom 25. März 2016) gilt sowohl für die Studie 190-201 als auch für die Studie 190-202. Mit Amendement 2 (07. März 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. Allerdings ist der finale SAP auf den 25. März 2016 datiert und somit vier Monate nach der letzten Visite in der Studie 190-201. Inwieweit möglicherweise Änderungen zur Datenauswertung in Kenntnis der bestehenden Daten festgelegt wurden, ist nicht nachzuvollziehen. So ist in der letzten Version des Studienprotokolls der Studie 190-201 (Amendement 5, vom 24. September 2014) ausgeführt, dass im SAP festgelegt sei, dass sowohl die Motor- und Language-Subskalen als auch die Gesamtskalen des Hamburg- und WCMC-Instruments mit den historischen Kontrolldaten verglichen würden. Dies ist im finalen SAP jedoch nicht enthalten. Änderungen in der Zielsetzung beinhalteten im SAP unter anderem: Immunogenität explorativer Endpunkt im Protokoll und sekundärer Endpunkt im SAP und QoL sekundärer Endpunkt im Protokoll und explorativer Endpunkt im SAP.

Im finalen SAP der Studie 190-201/190-202 war für die Primäranalyse ein Vergleich gegenüber einer historischen Kontrollgruppe vorgesehen. Auf Empfehlung der Zulassungsbehörden wurde im SAP der „Integrated Summary of Efficacy“ (SAP ISE, finale Version 3 vom 25.04.2016) 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 (siehe Kapitel 2.5.4.3).

2.5.4.1 Studie 190-201/190-202

Fallzahlplanung

Die Fallzahlschätzung erfolgte ausschließlich für die Studie 190-201. Eine Fallzahlschätzung für die Studie 190-202 war nicht erforderlich, da es sich um eine Extensionsstudie der Studie 190-201 handelt, an der alle Patienten, die die Studie 190-201 abgeschlossen hatten, teilnehmen konnten.

Der ursprüngliche primäre Endpunkt, auf dem die Fallzahlschätzung beruhte, war die intraindividuelle Schätzung der Steigung (‚Slope') für die Rate der Abnahme (‚rate of decline') der ML-Skala (0–6 Punkte) im Laufe der Zeit. Zur Fallzahlschätzung wurden Längsschnittdaten zum Krankheitsverlauf bei 30 unbehandelten Patienten mit genotypisch gesicherter CLN2 herangezogen. Danach 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 (µBMN-190) diesem Wert entspricht oder nicht:

- H_0: µBMN-190 = 2,0
- H_1: µBMN-190 ≠ 2,0

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. 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 190-201 aufgenommen (während der stabilen Dosisphase).

Zwischenauswertungen

Sowohl in der Studie 190-201 und 190-202 war geplant, die Studiendaten regelmäßig zur gemeinsamen Überprüfung durch BioMarin und das Datenüberwachungskomitee (DMC) deskriptiv zusammenzufassen. Ein frühzeitiges Stoppen erfolgte ausschließlich auf Basis der Sicherheitsdaten und des klinischen Urteils. Es wurden keine inferentiellen Stoppregeln verwendet. Zwischenanalysen zur Wirksamkeit und Sicherheit waren im September 2014 und Januar 2015 geplant.

Die Studie 190-202 ist noch nicht abgeschlossen. Die letzte Visite findet erwartungsgemäß am 15. Dezember 2020 statt. Für diese noch laufende Studie liegen zwei nicht a priori definierte Interimsanalysen vor: 15. Oktober 2015 und gemäß Modul 4 ein Datenschnitt vom 03. Juni 2016, der auf Ersuchen der FDA stattgefunden hat.

<u>Studienpopulationen</u>

Folgende Studienpopulationen wurden für die Wirksamkeitsanalysen im SAP für die Studien 190-201 und 190-202 definiert:

- *Intent-to-Treat-(ITT)-Population (N=23):* Alle Patienten, die die Studienmedikation erhalten haben, unabhängig von der Dosis, und bei denen zumindest ein Wirksamkeitsergebnis vorlag. Ein Patient der ursprünglich 24 eingeschlossenen Patienten 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 Dosiseskalationsphase erhalten hatte und der Abbruch nicht mit dem Behandlungsergebnis in Zusammenhang stand.
- *Per-Protocol-(PP)-Population (N=23):* Die PP-Population wurde zur Überprüfung der Robustheit als Sensitivitätsanalyse herangezogen. Sie umfasst alle Patienten der ITT-

Population, die den wesentlichen Protokollvorgaben entsprachen. Nicht berücksichtigt werden erhebliche Protokollabweichungen, die die Dateninterpretierbarkeit beeinflussen, und wiederholte übersprungene Dosen der Studienmedikation. Gründe für einen Ausschluss von Patienten waren vor der finalen Datenbanksperre zu definieren. Nur Patienten, die den wesentlichen Ein- und Ausschlusskriterien entsprachen, wurden in der PP-Population eingeschlossen:

- Einschlusskriterien: Alter ≥ 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).

  (Hinweis: Die Patienten der PP-Population wurden mindestens 23 Wochen mit einer Dosis von 300 mg behandelt und erhielten 80 % oder mehr der geplanten Infusionen während der 300-mg-Dosisphase).

  Hinweis: *Da die PP-Population mit der ITT-Population identisch ist, wurden in den Studien 190-201 und 190-202 keine zusätzlichen Analysen mittels PP-Population durchgeführt.*
- Safety-Population (N=24): Die Analyse zur Sicherheit umfasste alle Patienten, denen ein ICV-Port implantiert worden war.

Statistische Modelle und Tests

Analyse des primären Endpunkts (ML-Skala, Responderanalyse)

Im Studienverlauf wurde die Primäranalyse der Krankheitsprogression, gemessen mittels ML-Skala, von einer Slope- in eine Responder-Analyse geändert (SAP Version 3 vom 25.03.2016, lt. Angaben des pU aufgrund von Diskussionen mit den Zulassungsbehörden [FDA]). Gemäß Modul 4 des Dossiers blieb der ursprüngliche primäre Endpunkt ‚Slope' als Endpunkt erhalten. Die Analyse des primären Endpunkts (ML-Skala, Responderanalyse) ist in beiden Studien 190-201 und 190-202 identisch.

Definition der Response: Der primäre Endpunkt war der Anteil Patienten, die auf die Therapie angesprochen hatten (Responder; ITT). Responder waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 bzw. Woche 80 (Studientag 340 bzw. 540 bezogen auf die erste 300-mg-Infusion) nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme von maximal 1 Punkt). Demgegenüber waren Non-Responder (bzw. stabiler Abfall) definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 bzw. Woche 80 (Studientag 340 bzw. 540 bezogen auf die erste 300-mg-Infusion) um zwei oder mehr Punkte (≥ 2) abfiel. Ein so genannter stabiler Abfall entsprach demnach einer Verringerung um zwei oder mehr Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder um einen Punkt (oder mehr) zugenommen hatte.

Für die einarmige Studie 190-201 (ohne historischen Vergleich) wurde die Wirksamkeit von Cerliponase alfa mittels Einstichprobentest (exakter Binomialtest) wie folgt überprüft (pBMN-190 = Anteil Patienten in Prozent, die auf Cerliponase alfa angesprochen hatten):

- H_0: pBMN-190 ≤ 0,5
- H_1: pBMN-190 > 0,5

Der Wert von 0,5 wurde auf Basis von Längsschnittdaten (DEM-CHILD-Register) zum Krankheitsverlauf bei 30 unbehandelten Patienten mit genotypisch gesicherter CLN2 festgelegt, wonach der Score im Mittel um 2 Punkte/48 Wochen (SD 1) abnahm. Unter der Annahme, dass die Werte normalverteilt sind, nimmt der Wert erwartungsgemäß bei 50 % der unbehandelten Patienten um weniger als 2 Punkte ab und bei 16 % um weniger als 1 Punkt (H_0: pBMN-190 ≤ 0,25).

Analyse des ursprünglichen primären Endpunkts (mittlerer Abfall des ML-Scores, Slopeanalyse)

Zusätzlich wurde der primäre Endpunkt hinsichtlich der Abnahme des ML-Scores (Progressionsrate) skaliert auf 48 Wochen untersucht. Zur Analyse wurde der Einstichproben-t-Test eingesetzt und folgende Hypothese überprüft (µBMN-190 = die mittlere Abnahmerate der mit Cerliponase alfa behandelten Patienten):

- µBMN-190 ≥ 2,0
- µBMN-190 < 2,0

Analog der Responderanalyse basieren die Hypothesengrenzen auf Basis von Längsschnittdaten (DEM-CHILD-Register) zum Krankheitsverlauf bei 30 unbehandelten Patienten mit genotypisch gesicherter CLN2. Algorithmus zur Berechnung der Rate des Abfalls:

$$\text{Slope}=\frac{(Finaler\ ML{-}Score){-}(ML{-}Score\ zu\ Beginn)}{(Finaler\ Zeitpunkt){-}(Zeitpunkt\ zu\ Beginn)}$$

Sensitivitätsanalysen für die Responder- und Slopeanalyse (ML-Skala)

Sowohl für den primären Endpunkt (ML-Skala, Responderanalyse) als auch für den ursprünglichen Endpunkt (ML-Skala, Slope-Analyse) wurden a priori im finalen SAP folgende Sensitivitätsanalysen definiert:

- *Sensitivitätsanalyse I (N=21):* Ausschluss von Patienten mit einem Baseline-ML-Score von 6, der über den Verlauf der Studie unverändert blieb (Rationale laut pU: Patient sei noch nicht in die Phase der klinischen Verschlechterung [ML-Score < 6] eingetreten)
- *Sensitivitätsanalyse II (N=18):* Ausschluss von Patienten, deren Baseline-ML-Score von 6 über den Verlauf der Studie unverändert blieb oder deren Baseline-ML-Score < 3 Punkte war (Rationale laut pU: entsprach Einschlusskriterien, da trotz Screeningwerten von ≥ 3 Punkten einige Patienten bei Baseline bereits Punkte von < 3 aufwiesen)

Zusätzlich wurden im Studienbericht weitere Populationen für Sensitivitätsanalysen definiert, die allerdings für die vorliegende Nutzenbewertung nicht berücksichtigt werden, da ein Patient von den ursprünglich 24 eingeschlossenen Patienten in die Dosis-Eskalationsphase aufgenommen wurde, die Studie nach nur einer Infusionsgabe abgebrochen hat und somit kein 300-mg-Baselinewert für diesen Patienten vorliegt und zu Verzerrungen führen könnte.

- *Sensitivitätspopulation III (N=24):* Alle Patienten, die ihre informierte Zustimmung in die Studie 109-201 gegeben haben.

- *Sensitivitätspopulation IV (N=22):* Alle Patienten, die ihre informierte Zustimmung in die Studie 109-201 gegeben haben. Ausgeschlossen wurden Patienten mit einem stabilen ML-Score von 6.

Als weitere Sensitivitätsanalyse wurde die Responderanalyse für den primären Endpunkt wiederholt: Responder waren nun definiert als Patienten, die im Netto um weniger als 1 Punkt (< 1 Punkt) zu Woche 48 bzw. Woche 80 abfielen. Dieses wurde mittels Einstichprobentest (exakter Binomialtest) wie folgt getestet:

- H_0: pBMN-190 ≤ 0,25
- H_1: pBMN-190 > 0,25

Ereigniszeitanalysen zum primären Endpunkt

Die Analysen des primären Endpunktes werden mittels Kaplan-Meier dargestellt. Hierzu zählen: Zeit von Baseline bis zum ersten stabilen Abfall auf der ML-Skala (kein Wiederanstieg des ML-Scores bzw. Non-Responder) und die Zeit von Baseline bis zum ersten stabilen Abfall um 2 Punkte (kein Wiederanstieg des ML-Scores bzw. Non-Responder) oder Wert von Null auf der ML-Skala.

MLV- und MLVS-Skala

Eine Darstellung der MLV-Skala (0–9 Punkte, motorische/sprachliche Fähigkeiten plus Sehvermögen) und MLVS-Skala (0–12 Punkte, motorische/sprachliche Fähigkeiten/Sehvermögen plus Krampfaktivität) erfolgte deskriptiv als Veränderung zu Baseline.

Subgruppenanalysen

Alle Subgruppenanalysen erfolgten in den Studien 190-201/190-202 explorativ. Spezifische statistische Methoden waren nicht festgelegt. Homogenitäts- oder Interaktionstests waren nicht vorgesehen. Gemäß finalem SAP konnten folgende Subgruppen-Merkmale hinsichtlich einer möglichen Effektmodifikation (Wirksamkeit) untersucht werden:

- häufiger Genotyp (insbesondere c.622C>T- und c.509-1G>C-Mutationen),
- demographische Merkmale,
- Studienzentrum oder geographische Region,
- verschiedene Behandlungsintervalle (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).

Missing/Fehlende Werte

In beiden Studien 190-201/190-202 werden fehlende Werte als „missing“ gewertet, wenn nicht anders angegeben. Fehlende Werte der primären Responder-Analyse wurden durch den Einsatz der Kaplan-Meier-Methode adressiert. Hinsichtlich der Slope-Analysen spielten fehlende Werte keine nennenswerte Rolle, da zur Schätzung des Slope lediglich die Daten zu Baseline und zum letzten Erhebungszeitpunkt berücksichtigt wurden. Andere Imputationen waren nicht vorgesehen.

Baseline-Werte

Die Definition von Baseline variierte kontextabhängig und war in beiden Studien 190-201 und 190-202 identisch:

- *Safety-Baseline (I):* Zur Überprüfung der Sicherheit von Cerliponase alfa; Baseline umfasst die letzte Beobachtung vor der ersten Infusion von Cerliponase alfa.
- *Safety-Baseline (II):* Zur Überprüfung der Sicherheit von Cerliponase alfa, das die Sicherheit des ICV-Zugangs berücksichtigt; Baseline umfasst die letzte Beobachtung vor Implantation des ICV-Zugangs. Lag kein Wert vor der Implantation vor, entsprach Baseline der ersten Beobachtung nach der ICV-Implantation.

Wirksamkeit-Baseline:

- *Studien-Baseline:* Zur Überprüfung der Wirksamkeit von Cerliponase alfa in jeder Dosierung entsprach Baseline der letzten Beobachtung vor der ersten Infusion von Cerliponase alfa.
- *300-mg-Baseline:* Zur Überprüfung der Wirksamkeit von Cerliponase alfa in einer Dosis von 300 mg bezogen auf die ML-Skala und PedsQL™QoL-Bewertung entsprach Baseline der letzten Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet).

Die primären und sekundären Endpunkte wurden während der Dosiseskalation erhoben, die explorativen Endpunkte (QoL) allerdings nicht. Bei den Patienten, die an der Dosiseskalation 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. Für die Wirksamkeitsanalyse wurden die Patienten in der stabilen Dosisphase berücksichtigt und die Patienten, die bereits während der Dosis-Eskalationsphase 300 mg Cerliponase alfa erhielten. Bei der primären Wirksamkeitsanalyse sowie zur Analyse der Sicherheit wurden keine Zeitfenster eingesetzt.

Änderungen SAP

- *SAP Version 2.0 (09.12.2015):* 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 werden. 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.
- *SAP Version 3.0 (25.03.2016):* Nach Diskussionen mit den Zulassungsbehörden wurde eine Responder-Analyse als primärer Endpunkt definiert; die ursprüngliche Analyse blieb allerdings als wichtiger Endpunkt erhalten. Der primäre Endpunkt war definiert als ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion), der nicht im Netto um zwei Punkte (Responder: Abfall < 2) abfiel oder für den ein Wert von Null registriert wurde. Näher erläutert wurde der Aspekt des sogenannten stabilen Abfalls, eines Abfalls um zwei oder mehr Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder um einen Punkt (oder mehr) zugenommen hatte.

2.5.4.2 Studie 190-901

Auf Nachfrage bestätigt der pU am 31.07.2017, dass für die Studie 190-901 weder ein Protokoll noch ein statistischer Analyseplan zum Natural History Report vorliegt, da das DEM-CHILD-Register unabhängig von Seiten des pU aufgesetzt wurde. Die nachfolgenden Angaben sind dem Studienbericht entnommen und werden zur Vollständigkeit berichtet, allerdings für die vorliegende Nutzenbewertung nicht berücksichtigt, da der SAP der „Integrated Summary of Efficacy“ (SAP ISE, finale Version 3 vom 25.04.2016) für den Wirksamkeitsvergleich nach 1:1-Paarbildung der Teilnehmer der Studie 190-201/202 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) im Fokus steht.

Zur Kalkulation der klinischen Krankheitsprogression (Abfall des HML-Skalenwerts) über 48 Wochen (1 Monat = 30 Tage) wurden zwei Algorithmen eingesetzt:

- Erster/Letzter Wert: Die Linie zwischen folgenden zwei HML-Skalenwerten wurde als Rate (Slope) der Krankheitsprogression verwendet. Der erste Wert entsprach dem zeitlich ersten HML-Skalenwert unter 6. Falls mehrere gleiche Werte unter 6 vorlagen, wurde der mittlere Zeitpunkt dieser Werte berücksichtigt. Der letzte Wert entsprach dem letzten Score > 0.
- Regression: Als Rate (Slope) der Krankheitsprogression wurde die mittels linearer Regression geschätzte Linie zwischen dem ersten und letzten HML-Skalenwert wie oben und allen weiteren dazwischenliegenden HML-Scores verwendet.

2.5.4.3 Historischer Vergleich der Wirksamkeitsdaten (ML-/HML-Skala) von Studie 190-201/190-202 vs. Studie 190-901 (Integrated Summary of Efficacy, ISE)

Der ursprüngliche SAP der Studie 190-201/190-202 sieht für den primären Endpunkt einen Vergleich gegenüber einer aus den historischen Verlaufsdaten festgesetzten historischen Kontrollgruppe vor. Dieser wurde erst auf Empfehlung der Zulassungsbehörden im SAP ISE konkretisiert, wonach 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 wurde. Der SAP ISE liegt in der finalen 3. Version vom 25. April 2016 vor.

Studienpopulationen

Tabelle 9: Studienpopulationen für den Vergleich

	Studie 190-201	Studie 190-901
ITT-Population	alle Patienten der Studie 190-201, die die Studienmedikation erhalten haben und Wirksamkeitsergebnisse berichten[1)]	Patienten mit: Alter ≥ 3 Jahre und mindestens ein CLN2-Score ≥ 3 im Alter ≥ 3 Jahren (entsprechend den Einschlusskriterien von Studie 190-201)
Population für Sensitivitätsanalyse[2)]	I.) Ausschluss von Patienten mit einem ML-Score von 6 zu Baseline, der über den Verlauf der Studie unverändert blieb	
	II.) Ausschluss von Patienten, deren Baseline-ML-Score von 6 über den Verlauf der Studie unverändert blieb oder deren Baseline-ML-Score < 3 Punkte war (Rationale laut pU: entsprach Einschlusskriterien, da trotz Screeningwerten von ≥ 3 Punkten einige Patienten bei Baseline bereits Punkte von > 3 aufwiesen)[3)]	

	Studie 190-201	Studie 190-901
Population für die primäre Wirksamkeitsanalyse	Methode #1: 1:1-Matching auf Basis eines identischen ML-/HML-Wertes und einem Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline; berücksichtigte Erhebungszeitpunkte der Nachbeobachtung des gematchten Patienten aus Registerstudie 901: so lange wie möglich (falls nur < 6 Monate, wird dieser Matchingpartner nicht herangezogen), aber nicht länger als das vollständige Follow-up der gepaarten Patienten der Studie 190-201.	
Population für supportive Analysen	Methode #1 (Simulation): 1.000 Simulationen wurden 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 ≤ 12 Monate). Die Responder-Analyse wurde für jede der 1.000 Simulationen wiederholt.	
	Methode #2:1:1-Matching anhand eines zweiten Auswahlalgorithmus (1:1 second algorithm): Quadratsumme des Abstands zwischen den Baseline-ML-Scores und dem Baseline-Alter (in Monaten) in beiden Studien ist ≤ 36 und mittleres Abstandsquadrat ist möglichst niedrig. Somit 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 Nachbeobachtung entsprach der Methode #1.	
	Methode #3 (Many-to-one-Paarbildung)[3]: Paarbildung (jeweils zu Baseline) anhand der Variablen ML-Score, Alter und häufiger Genotyp. Folgende drei Matching-Sets wurden gebildet: • ML/HML-Score zu Baseline, • ML/HML-Score und das Alter jeweils zu Baseline und • ML/HML-Score zu Baseline und Genotyp (c.622>T und/oder c.509-1G>C vorhanden oder nicht). • Für jeden Patienten aus Studie 201/202 konnten sich damit 0, 1 oder mehrere Matching-Partner ergeben. Verglichen mit der 1:1-Paarbildung ermöglicht das Matching den Einschluss einer größeren Zahl Patienten der Studie 190-901.	
	Methode #4 (Non-Matching): Unter Verwendung der ITT-Population der Studien 190-201 und 190-901	

[1] Ein Patient der ursprünglich 24 eingeschlossenen Patienten brach die Studie nach einer einzigen Infusion ab. Dies erfolgte auf Wunsch der Eltern. Dieser Patient wurde aus der ITT-Population ausgeschlossen.
[2] Rationale für die Wahl der Sensitivitätsanalysen (siehe Kapitel 2.5.4.1)
[3] Diese Sensitivitätsanalyse und Methode #3 (Many-to-one-Paarbildung) wurden im finalen ISE SAP präspezifiziert. Allerdings finden sich hierzu im Studienbericht (Datenschnitt: 15. Juni 2016) für die Endpunkte der (H)ML-Skalen keine Analysen bzw. Ergebnisse.

Abkürzungen: CLN2: Ceroid-Lipofuszinose Typ 2; ITT: Intention-to-treat

Baseline-Werte

Baselinewert der Studien 190-201 und 190-202 ist die letzte Beobachtung vor der ersten 300-mg-Infusion von Cerliponase alfa. Für den Vergleich mit der historischen Kontrollstudie 190-901 wurden folgende Baselinewerte definiert:

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.

Missing/Fehlende Werte

- Weder die Analysen der Krankheitsprogression (Slopes) noch die Ereigniszeitanalysen erforderten eine Imputation fehlender Werte.
- Die Erhebungszeitpunkte des ML-/HML-Scores in den Studien 190-201/190-202 und 190-901 unterschieden sich. 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:
  - Vorzugsweise wurde eine lineare Interpolation eingesetzt. Lag lediglich ein Einzelwert innerhalb von 28 Tagen bezogen auf den Analysetag vor, wurde dieser berücksichtigt.
  - Eingabe desjenigen Werts, der dem Zeitpunkt am nächsten lag („windowing“).
- Der ursprüngliche Datensatz enthielt abgeleitete (imputierte) Visiten, die identifiziert und entfernt wurden.
- 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 HML-Wert eingegeben:
  - In der Studie 190-901 fanden keine regulären Erhebungen der ML-Skala statt und die Erhebungszeitpunkte unterschieden sich somit von denen der Studie 190-201. Um eine Analyse beider Patientengruppen zu bestimmten Zeitpunkten zu ermöglichen, wurden die (H)ML-Ergebnisse der Patienten beider Studien zu den folgenden nominalen Zeitpunkten imputiert. Die nominalen Zeitpunkte waren Woche 25, 37, 49 und 73 (entsprechend Tag 169, 253, 337 und 505; Woche 73: Studie 190-202). Bei Patienten

der Studie 190-901, bei denen sich der Score zwischen zwei Visiten nicht änderte, wurde dieser Wert monatlich eingegeben. Hierfür wurde auf Basis der linearen Interpolation geschätzt: Lagen keine Werte sowohl vor als auch nach dem Analysetag vor, jedoch ein Einzelwert innerhalb von 28 Tagen, dann wurde der Einzelwert eingegeben.

Zwischenauswertungen/vorliegende Datenschnitte

Für den Vergleich liegen drei nicht a priori definierte Interimsanalysen vor: 30. November 2015; 15. Juni 2016 und 01. November 2016. Die letzten beiden Datenschnitte wurden auf Ersuchen der FDA durchgeführt.

Statistische Modelle und Tests

Analyse des primären Endpunkts (ML-/HML-Skala, ISE Responderanalyse)

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 ML-/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 Definition des primären Endpunkts Response der ISE differiert von der Responderdefinition in der einarmigen Studie 190-201/202: Die Response für den historischen Vergleich der ISE ist definiert anhand des Slope, d. h. anhand der Größe des Abfalls der motorischen und sprachlichen Fähigkeiten, gemessen mittels der Verlaufsdaten der (H)ML-Werte. Der lineare Abfall wurde anhand eines jeweiligen Startwertes, (H)ML-Wert zu Baseline, und eines Endwertes, letzter verfügbarer Messwert > 0, bestimmt, indem diese beiden Punkte zu einer Linie verbunden wurden. Responder waren somit definiert als Patienten, bei denen der Abfall (Slope der Linie), skaliert auf 48 Wochen, kleiner als zwei (< 2) war (festgelegt im finalen SAP ISE, 3. Version vom 25. April 2016)

- H_0: p190-201/202 = p190-901
- H_1: p190-201/202 ≠ p190-901,

wobei p190-201/202 = Anteil der Responder in Prozent, die auf Cerliponase alfa angesprochen hatten; p190-901 = Anteil der Responder in Prozent, die in Studie 190-901 angesprochen hatten.

Analyse des ursprünglichen primären Endpunkts (Krankheitsprogression [Abfall des ML-/HML-Scores], Slopeanalyse)

Unter Einsatz eines Zweistichproben-t-Tests wurde der mittlere Abfall bei behandelten und unbehandelten Patienten verglichen (zweiseitig auf einem Signifikanzniveau von p=0,05). Die Analyse erfolgte für die Matching Methode #1 (1:1) und wurde getrennt nach Domäne (motorische und sprachliche Fähigkeiten) analysiert. Darüber hinaus werden weitere Sensitivitätsanalysen mit anderen Matching-Methoden bzw. Non-Matching-Methoden durchgeführt (siehe Tabelle 9). Folgende Hypothesen wurden überprüft:

- H_0: μ190-201/202 = μ190-901
- H_1: μ190-201/202 ≠ μ190-901

wobei μ die mittlere Abnahmerate der Studienpopulation ist. Zusätzlich wurden die Analysen mit Hilfe der supportiven Population (Sensitivitätsanlysen) durchgeführt (siehe Tabelle 9). Zusätzlich wurde getrennt nach Domäne (motorische und sprachliche Fähigkeiten) analysiert.

Sensitivitätsanalysen für die Responder- und Slopeanalyse (ML-/HML-Skala)

Die Sensitivitätsanalysen entsprechen den a priori im SAP und Studienbericht der Studien 190-201/190-202 definierten Analysen. Zusätzlich wurden die Populationen als supportive Analysen für den primären Endpunkt herangezogen (siehe Tabelle 9). Darüber hinaus wurde der primäre Endpunkt getrennt nach den Domänen (motorische und sprachliche Fähigkeiten) analysiert; definiert als Anteil der Patienten, bei denen der ML-/HML-Score skaliert auf 48 um < 1 Punkt abnimmt.

Ereigniszeitanalysen für Endpunkte Response stabiler Abfall des ML-/HML-Scores

Darüber hinaus wurden im ISE SAP Ereigniszeitanalysen von Baseline bis zur letzten Erhebung mittels Kaplan-Meier (und eines Cox-Modells, adjustiert für ML-/HML-Score und Alter zu Baseline) für folgenden Endpunkt definiert: Response (definiert als Nichtauftreten eines stabilen Abfalls um 2 Punkte, dem kein Wiederanstieg des ML-/HML-Scores folgt, oder als Nichtauftreten des Wertes von 0 der ML-/HML-Skala). Zusätzlich zu diesem Response-Endpunkt wurden mittels Kaplan-Meier analysiert: Zeit von Baseline bis zum ersten stabilen Abfall auf der ML-/HML-Skala, Zeit von Baseline bis zum ersten stabilen Abfall um ≥ 2 Punkte auf der ML-/HML-Skala (jeweils ohne Wiederanstieg des ML-/HML-Scores) oder Zeit bis zu einem Wert von Null.

MLV- und MLVS-Skala

Eine Darstellung der MLV-Skala (0–9 Punkte, motorische/sprachliche Fähigkeiten plus Sehvermögen) und MLVS-Skala (0–12 Punkte, motorische/sprachliche Fähigkeiten/Sehvermögen plus Krampfaktivität) erfolgte deskriptiv als Veränderung zu Baseline.

2.6 Ergebnisse zum Zusatznutzen

Nachfolgend werden die Ergebnisse der Studien 190-201/190-202 sowie die Ergebnisse der ISE-Analyse, also des Vergleichs nach 1:1-Paarbildung der Studienteilnehmer der Studie 190-201/190-202 mit den Patienten der historischen Kontrollgruppe 190-901, dargestellt.

Die berichteten Ergebnisse der Studie 190-201 bilden den Datenschnitt vom 30. November 2015 (Studienende) ab. Zu diesem Zeitpunkt hatten alle Patienten die 48 Wochen Behandlung mit Cerliponase alfa (300 mg) abgeschlossen. Für die Extensionsstudie 190-202 wird der aktuelle Datenschnitt vom 03. Juni 2016 herangezogen. Zu diesem Zeitpunkt betrug die Behandlungsdauer von 300 mg Cerliponase alfa bei 96 % der eingeschlossenen Patienten mehr als 72 Wochen.

Für den historischen Vergleich wird der Datenschnitt vom 15. Juni 2016 herangezogen, wonach alle Patienten über mindestens 72 Wochen mit Cerliponase alfa behandelt wurden.

2.6.1 Charakterisierung der Studienpopulation

Studie 190-201/190-202 (Datenschnitt: 3. Juni 2016)

Insgesamt erfüllten 24 Patienten die Einschlusskriterien und wurden in die Studie 190-201 eingeschlossen. Ein Patient der Kohorte 3 in der Dosis-Eskalationsphase (300 mg) zog seine Einverständniserklärung nach einer Dosis des Prüfpräparates zurück. Vierzehn Patienten begannen direkt mit der stabilen Dosisphase (300 mg/ alle 2 Wochen) für 48 Wochen. Insgesamt beendeten 23 Patienten die 48 Wochen der Studie 190-201 und wurden in die Studie 190-202 aufgenommen. Zum Datenschnitt der ersten Zwischenanalyse der Studie 190-202 (15. Oktober 2015) hatten fünf von 23 Patienten die Studie 190-201 noch nicht abgeschlossen. Zum Zeitpunkt der nicht a priori definierten Zwischenanalyse vom 3. Juni 2016 nahmen alle 23 Patienten an der Studie teil. Wie eingangs beschrieben, beziehen sich die nachfolgend berichteten Ergebnisse auf den aktuelleren Datenschnitt vom 03. Juni 2016, welcher alle Daten der Studie 190-201 sowie der nachfolgenden Studie 190-202 bis zum Datenschnitt 3. Juni 2016 umfasst (siehe Tabelle 10).

Während der gesamten Studie 190-201 wurden insgesamt 616 Infusionen verabreicht, wovon 587 Infusionen ausschließlich die stabile Dosis von 300 mg umfassten. Eine Infusion wurde während der gesamten Studie bei einem Patienten verpasst. Von den verabreichten 616 Infusionen wurden 16 Infusionen bei neun Patienten unterbrochen und acht Infusionen bei sechs Patienten nicht abgeschlossen (siehe Tabelle 10).

Tabelle 10: Allgemeine Angaben zur Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016)

Studie 190-201/190-202	**Studie 190-201/190-202** n (%)
in die Studie eingeschlossen Studie 190-201 Studie 190-202	 24 (100) 23 (96)
Studie beendet Studie 190-201 Studie 190-202[1)] *(geplant 15. Dezember 2020)*	 23 (96) 0
Abbruch der Einnahme der Studienmedikation Entzug der Einverständniserklärung und Abbruch der Studie 190-201[2)]	 1 (4)
ITT-Population/PP-Population[3)]	23 (96)
Safety-Population[4)]	24 (100)
Sensitivitätsanalyse I[5)]	21 (88)
Sensitivitätsanalyse II[6)]	18 (75)
Anzahl Infusionen (insg.) auf Patientenbasis in der Studie 190-201 (Datenschnitt: 30. November 2015)[7)] unabhängig von der Dosis *verpasste/fehlende Infusionen (bezogen auf N=24)* *nicht abgeschlossene Infusionen[8)] (bezogen auf N=24)* *unterbrochene Infusionen[9)] (bezogen auf N=24)* stabile Dosis (300 mg)	 616 6 (25) 9 (38) 16 (67) 587
Mediane Anzahl Infusionen auf Patientenbasis in der Studie 190-201 (Datenschnitt: 30. November 2015)[7)] (min;max) 30 mg 100 mg 300 mg	 4 (2;6) 2,5 (2;5) 24 (1;31)

Studie 190-201/190-202	Studie 190-201/190-202 n (%)
unabhängig von der Dosis	24 (1;35)
Anzahl Infusionen (insg.) auf Patientenbasis in der Studie 190-201/190-202 (Datenschnitt: 15.10.2015)[11)]	
unabhängig von der Dosis	805
verpasste/fehlende Infusion (bezogen auf N=24)	1 (4)
nicht abgeschlossene Infusionen[8)] (bezogen auf N=24)	8 (33)
unterbrochene Infusionen[9)] (bezogen auf N=24)	13 (54)
300 mg	776
Mediane Anzahl Infusionen in der Studie 190-201/190-202[7)] (min;max)	
30 mg	4,0 (2;6)
100 mg	2,5 (2;5)
300 mg	48,5 (1;63)
unabhängig von der Dosis	48,5 (1;71)
Mediane Exposition[10)] in der Studie 190-201/190-202 in Wochen[7)] (min;max)	
30 mg	6,4 (1,7;10;7)
100 mg	3,2 (2,1;8,1)
300 mg	95,1 (0,1;124,4)
unabhängig von der Dosis	95,1 (0,1;141)

[1)] Studie 190-202 wird erst im Dezember 2020 abgeschlossen sein. Zum aktuellen Datenschnitt (3. Juni 2016) befinden sich alle 23 Patienten in der Extensionsstudie 190-202.
[2)] Ein Patient der Studie 190-201 zog seine Einverständniserklärung nach einer Infusion zurück.
[3)] Alle Patienten, die die Studienmedikation erhalten haben, unabhängig von der Dosis, und bei denen zumindest ein Wirksamkeitsergebnis vorlag. Ein Patient brach die Studie nach nur einer Infusion ab. Dieser Patient wurde aus allen Analysen einschließlich ITT ausgeschlossen. Da der klinische Review keine Hinweise auf bedeutende Protokollverletzungen lieferte, waren die ITT- und PP-Population identisch.
[4)] Umfasst alle Patienten, denen ein ICV-Port implantiert worden war.
[5)] Sensitivitätsanalyse I: ITT, ausgeschlossen sind Patienten, die zu Baseline einen ML-Score von 6 hatten und keine Abnahme des ML-Scores über den Studienverlauf berichteten.
[6)] Sensitivitätsanalyse II: ITT, ausgeschlossen sind Patienten, die zu Baseline einen ML-Score von 6 hatten und keine Abnahme des ML-Scores über den Studienverlauf berichteten und deren ML-Score zu 300-mg-Baseline < 3 Punkte war.
[7)] Safety-Population
[8)] Infusion wurde vor 10 ml gestoppt und nicht weitergegeben.
[9)] Infusion wurde vorzeitig gestoppt, unabhängig davon, ob diese wieder weiter gegeben wurde oder nicht.
[10)] Exposition in mg = Infusionsvolumen (ml) × Wirkstoffkonzentration (mg/ml)
[11)] 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. Die anderen fünf Patienten wurden am 01. Dezember 2015 in die Studie eingeschlossen. Keine aktuellen Angaben im Zwischenbericht vom 3. Juni 2016.

Abkürzungen: ITT: Intention-to-treat; mg: Milligramm; max: Maximum; min: Minimum; PP: Per Protokoll

Die soziodemographischen und krankheitsbezogenen Charakteristika der Patienten der Studie 190-201 sind in Tabelle 11 dargestellt und beziehen sich auf den Datenschnitt vom 15. Oktober 2015, da im Datenschnitt vom 03. Juni 2016 keine Daten vorhanden sind.

Das mediane Alter bei Studienschluss lag bei 4 Jahren (Spanne: 3 bis 8 Jahre). Die Diagnosestellung lag im Median bei 3 Jahren (Spanne: 2,5 bis 6,3 Jahre). Die Mehrheit der Studienteilnehmer war weiblich (63 %) und Kaukasier (96 %). Es wurden gemäß Studienprotokoll ausschließlich Patienten mit einem Gesamtscore von 3–6 auf der ML-Skala eingeschlossen. Insgesamt 20 von 24 Patienten hatten zur Screening-Visite einen ML-Score von 3–4 Punkten und entsprechend einen ML-Score im unteren protokollgemäßen Bereich. 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 mittlere ML-

Score 3,5 Punkte (SD 1,2). Zu diesem Zeitpunkt hatten drei Patienten einen ML-Sore unter drei (n=1 ML-Score von 1; n=2 ML-Score von 2). Alle Patienten hatten mindestens eine Vorerkrankung. Darüber hinaus erhielten alle Patienten der Studie 190-201 (n=24, 100 %) vor der ersten Infusion bereits mindestens eine Vorbehandlung mit: Propofol (88 %), Midazolam (67 %), Paracetamol (50 %), Cefuroxim (46 %), Sufentanil (46 %), Levetiracetam (42 %) und Natriumvalproat (42 %).

Tabelle 11: Charakterisierung der Studienpopulation der Studie 190-201/190-202 (eingeschlossene Patienten [N=24], Datenschnitt: 15.10.2015[7])

Studie 190-201/190-202	**Studie 190-201/190-202 N=24**
Alter (Jahre) bei Studieneinschluss	
Mittelwert (SD)	4,3 (1,24)
Median (min;max)	4 (3;8)
Alter (Jahre) bei Diagnosestellung	
n	23[6]
Mittelwert (SD)	3,4 (1,07)
Median (min;max)	3 (2,5;6,3)
< 3 Jahre	7 (29)
3 bis < 5 Jahre	12 (50)
≥ 5 Jahre	4 (17)
präsymptomatisch	1 (4)
Geschlecht, n (%)	
Männlich	9 (38)
Weiblich	15 (63)
Ethnie (genetisch), n (%)	
Kaukasier/weiß	23 (96)
asiatisch	1 (4)
Region, n (%)	
Hamburg	12 (50)
London	2 (8)
Columbus	3 (13)
London	1 (4)
Italien	6 (25)
Genotyp, n (%)	
c.622C>T	5 (21)
c.509-1G>C	2 (8)
c.622C>T und c.509-1G>C	2 (8)
c.622C>T und andere	4 (17)
c.509-1G>C und andere	4 (17)
andere	7 (29)
fehlend	-
ML-Score zu Screening (min;max)	
Mittelwert (SD)	3,7 (0,95)
Median (min;max)	3 (3;6)
6	2 (8)
5	2 (8)
4	7 (29)
3	13 (54)

Studie 190-201/190-202	Studie 190-201/190-202 N=24
ML-Score zu Studien-Baseline[1)] (min;max)	
Mittelwert (SD)	3,6 (1,06)
Median (min;max)	3 (2;6)
6	2 (8)
5	2 (8)
4	6 (25)
3	12 (50)
2	2 (8)
ML-Score zu 300-mg-Baseline[2)] (min;max)	
Mittelwert (SD)	3,5 (1,2)
Median (min;max)	3 (1;6)
6	2 (8)
5	2 (8)
4	6 (25)
3	11 (46)
2	2 (8)
1	1 (4)
Hamburg-Gesamtskala (MLVS)[3)] zum Screening bzw. Baseline (min;max)	
Mittelwert (SD)	8 (1,81)
Median (min;max)	8 (4;12)
10–12	4 (16)
7–9	15 (63)
4–6	5 (21)
0–3	0
Patienten mit mindestens einer Vorerkrankung der Studie 190-201[4)]	24 (100)
(> 10 %) nach PT	
NCL	24 (100)
Obstipation	3 (13)
Dysphagie	3 (13)
Hypersalivation	3 (13)
Gangstörung	3 (13)
Ataxie	14 (58)
Dysarthrie	6 (25)
Epilepsie	18 (75)
Hypotonie	4 (17)
Sprachstörung	11 (46)
Motorische Funktionsstörungen	3 (13)
Myoklonus	3 (13)
Psychomotorische Fähigkeiten eingeschränkt	10 (42)
Krampfanfall	5 (21)
Sprachstörung entwicklungsbedingt	5 (21)
Patienten mit mindestens einer Vorerkrankung der Studie 190-202[5)]	24 (100)
(> 10 %) nach PT	
NCL	24 (100)
Obstipation	6 (25)
Dysphagie	5 (21)
Hypersalivation	4 (17)
Entwicklungsverzögerung	5 (21)

Studie 190-201/190-202	Studie 190-201/190-202 N=24
Gangstörung	6 (25)
Ataxie	15 (63)
Dysarthrie	6 (25)
Dystonie	4 (17)
Epilepsie	19 (79)
Plantar reflex	3 (13)
Hypotonie	5 (21)
Sprachstörung	14 (58)
Motorische Funktionsstörungen	3 (13)
Myoklonus	6 (25)
Psychomotorische Fähigkeiten eingeschränkt	10 (42)
Krampfanfall	6 (25)
Sprachstörungen	10 (42)
Sprachstörung entwicklungsbedingt	5 (21)
Tremor	4 (17)
Reizbarkeit	3 (13)

[1] Baseline entspricht der letzten Beobachtung vor der ersten Infusion (N=24).
[2] Baseline entspricht der letzten Beobachtung vor der ersten 300-mg-Infusion (N=23). Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis für Patienten der Dosis-Eskalationsphase nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet.
[3] Die Hamburg-Gesamtskala (0–12 Punkte), bestehend aus den vier Domänen motorische und Sprachkompetenz, Sehvermögen und Krampfanfälle wurde ebenfalls als Vollversion erhoben. Datenschnitt: 15. Oktober 2015
[4] Wurde während der Screeningphase der Studie 190-201 erhoben und dokumentiert.
[5] Wurde während der ersten Woche und vor der ersten Dosis in der Studie 190-202 erhoben. Andauernde UEs/SUEs, die bereits in der Studie 190-201 aufgetreten waren, wurden als Vorerkrankung in der Studie 190-202 dokumentiert.
[6] Die Charakterisierung der Studienpopulation bezieht sich auf die eingeschlossenen Patienten in die Studie 190-201 (N=24). Dennoch bezieht sich das ‚Alter bei Diagnosestellung' auf 23 Patienten. Darüber hinaus ist unklar, was einerseits unter präsymptomatisch zu verstehen ist, insbesondere vor dem Hintergrund der a priori definierten Einschlusskriterien (siehe Kapitel 2.4.1) N=23 Patienten bezieht und nicht auf N=24 (eingeschlossene Patienten) und andererseits, ob der eine präsymptomatische Patient in den Altersklassen (< 3 Jahre; 3 bis < 5 Jahre; ≥ 5 Jahre) mit umfasst ist.
[7] 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. Die anderen fünf Patienten wurden am 01. Dezember 2015 in die Studie eingeschlossen. Keine aktuellen Angaben im Zwischenbericht vom 3. Juni 2016.

Abkürzungen: ML: Motor/Language (Kombinierte Skala Motorik/Sprache)

Begleitmedikationen in der Studie 190-201/190-202

Tabelle 12 gibt einen Überblick über jene Begleitmedikationen, die während der Studien 190-201 (Datenschnitt: 30. November 2015) und 190-202 (Datenschnitt: 15. Oktober 2015) von 20 % und mehr der Patienten eingenommen wurden. In beiden Studien erhielten alle Patienten (N=24) mindestens eine Begleitmedikation. Dabei stellten Benzodiazepin-Derivate (100 %), andere Antiepileptika (83 %), Fettsäurederivate (88 %), Nicht-Opiod-Schmerz-/Fiebermittel wie Paracetamol (Studie 190-201: 92 %; Studie 190-202: 96 %), Glucocorticoide (79 %), Piperazinderivate (Studie 190-201: 63 %; Studie 190-202: 71 %) und andere Allgemeinanästhetika (79 %) die Medikamentengruppen dar, die am häufigsten eingenommen wurden.

Tabelle 12: Begleitmedikationen (≥ 20 %) der Studie 190-201/190-202 (eingeschlossene Patienten [N=24], Datenschnitt: 15. Oktober 2015)

ATC-Klasse, Preferred Term[1)]	**Studie 190-201/190-202[4)]** *n (%)*
Aldehyde und Derivate	**8 (33)**
Chloralhydrat	8 (33)
Aminoalkylether[2)]	**6 (25)**
Anilide	**23 (96)**
Paracetamol	23 (96)
Propacetamol Hydrochlorid	5 (21)
Antibiotika[2)]	**5 (21)**
Barbiturate und Derivate	**5 (21)**
Phenobarbital	5 (21)
Benzodiazepine Derivate	**24 (100)**
Clobazam	9 (38)
Clonazepam	6 (25)
Diazepam	17 (71)
Lorazepam	9 (38)
Midazolam	18 (75)
Herztherapie	**6 (25)**
Atropine	6 (25)
Kombination aus Penicillin inkl. Beta-Lactamase Inhibitoren[3)]	**6 (25)**
Klistier[3)]	**5 (21)**
Fettsäure Derivate	**21 (88)**
Natriumvalproat	18 (75)
Valproinsäure	11 (46)
Glucocorticoide	**19 (79)**
Betamethasone Sodium Phosphate	6 (25)
Prednisolon	12 (50)
Halogenkohlenwasserstoffe	**11 (46)**
Sevofluran	11 (46)
Melatonin-Rezeptor-Agonisten	**7 (29)**
Melatonin	7 (29)
Andere dermatologische Präparate	**10 (42)**
Andere dermatologische Präparate (undefiniert)	10 (42)
Opioide Anästhetika	**13 (54)**
Alfentanil	9 (38)
Alfentanil hydrochlorid	5 (21)
Osmotisch wirkende Abführmittel[2)]	**8 (33)**
Andere Antiepileptika	**20 (83)**
Lamotrigin	7 (29)
Levetiracetam	12 (50)
Zonisamid	6 (25)

ATC-Klasse, Preferred Term[1)]	Studie 190-201/190-202[4)] *n (%)*
Andere Anästhetika Propofol	**19 (79)** 19 (79)
Andere quaternäre Ammonium-Verbindungen Mivacurium	**8 (33)** 6 (25)
Piperazin Derivate Cetirizin hydrochlorid	**17 (71)** 16 (67)
Propionsäure-Derivate Ibuprofen	**16 (67)** 16 (67)
Pyrazolone[2)]	**5 (21)**
Cephalosporine der zweiten Generation Cefuroxim	**6 (25)** 6 (25)
Serotoninantagonisten Ondansetron	**6 (25)** 6 (25)
Lösungen, die den Elektrolythaushalt beeinflussen Dextrose- und Natriumchlorid-Injektionen (Glucose, Natriumchlorid)	**13 (54)** 10 (42)
Substituierte Alkylamine Chlorphenaminmaleat	**6 (25)** 6 (25)
Cephalosporine der dritten Generation	**6 (25)**

[1)] Patienten wurden nur einmal in jeder ATC-Klasse und Preferred Term gezählt.
[2)] Kein Preferred Term wurde mit ≥ 15 % der Patienten beobachtet.
[3)] CSR 27.04 2016 (Studie am 30. November 2015 beendet)
[4)] 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. Die anderen fünf Patienten wurden am 01. Dezember 2015 in die Studie eingeschlossen. Keine aktuellen Angaben im Zwischenbericht vom 3. Juni 2016.

Abkürzungen: ATC: Anatomisch-Therapeutisch-Chemisch; PT: Preferred Term

Studie 190-901 – Historische Kontrollgruppe (Datenschnitt: 1. Mai 2016)

Die Analyse auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901) dient als historische Kontrolle der Studien 190-201/190-202. Bei dieser Analyse wurden spezifische Auswahlkriterien eingesetzt, um eine Patientenpopulation zu erhalten, die den wesentlichen Ein- und Ausschlusskriterien der Studie 190-201 entsprach. Hierfür wurden verschiedene Kriterien angewendet. Von den 74 Patienten aus dem DEM-CHILD-Register hatten 58 Patienten mindestens einen berichteten HML-Score. Darüber hinaus wurden für die 58 verbliebenen Patienten zusätzliche Filter eingesetzt (nähere Informationen 2.4.2). Nach Einsatz der Kriterien waren 41 von 74 Patienten (55 %; auswertbare Patienten) in der Analyse und erfüllten die Eignungskriterien für den historischen Vergleich mit den Patienten der Studie 190-201/190-202 (siehe Tabelle 13).

Informationen zu früheren Therapien, Begleitmedikationen oder Vorerkrankungen liegen für die historische Kontrollgruppe nicht vor. Die Mehrzahl der eingeschlossenen Patienten war männlich (59 %) und die Diagnosestellung erfolgte im Median im Alter von 57 Monaten (Spanne: 35–118 Monate). Der mediane Diagnosezeitpunkt der 41 möglichen Matchingpartner der historischen Vergleichskohorte liegt damit ca. 21 Monate später als der mediane Diagnosezeitpunkt der Studie

190-201 (ca. 36 Monate). Erste klinische Anzeichen traten im Median bei 36 Monaten (Spanne: 18–24 Monate) auf. Die häufigsten klinischen Anzeichen waren Krampfanfälle (56 %) und Sprachauffälligkeiten (49 %). Dabei führten die Symptome Krampfanfälle (73 %), Sprachschwierigkeiten (59 %) und Ataxie (56 %) zur Diagnose. Die Verteilung der ML-/HML-Scores zu Baseline der Patienten aus der Studie 901-201 und zum Diagnosezeitpunkt der 41 möglichen Matchingpartner der historischen Vergleichskohorte ist unterschiedlich, mit tendenziell höherem Anteil von stärkerer Beeinträchtigung bei Patienten aus dem Kontrollregister (siehe Tabelle 10 und Tabelle 13).

Tabelle 13: Charakterisierung der Studienpopulation der Studie 190-901 (auswertbare Population [N=41], Datenschnitt: 01. Mai 2016)

Studie 190-901	**Hamburg N=33** *n (%)*	**Verona N=8** *n (%)*	**Insgesamt N=41** *n (%)*
Geburtsjahr			
1960–1969	1 (3)	0	1 (2)
1970–1979	1 (3)	2 (25)	3 (7)
1980–1989	2 (6)	0	2 (5)
1990–1999	16 (48)	3 (38)	19 (46)
2000–2009	13 (39)	3 (38)	16 (39)
Alter (Monate) bei erstem klinischen Anzeichen			
n	24	8	32
Mittelwert (SD)	34,6 (8,6)	39,3 (9,6)	35,8 (8,94)
Median (min;max)	35,5 (18;53)	42 (24;54)	36 (18;54)
Erste klinische Anzeichen			
Verhaltensauffälligkeiten	5 (15)	3 (38)	8 (20)
Demenz	1 (3)	0	1 (2)
Sprachschwierigkeiten	16 (48)	4 (50)	20 (49)
Lernschwierigkeiten	2 (6)	0	2 (5)
Motorische Schwierigkeiten	11 (33)	4 (50)	15 (37)
Krampfanfälle	17 (52)1)	6 (75)	23 (56)
Verlust des Sehvermögens	1 (3)	0	1 (2)
Nicht berichtet	7 (21)	0	7 (17)
unbekannt	1 (3)	0	1 (2)
Alter (Monate) beim ersten Auftreten von Krampfanfällen			
n	16	6	22
Mittelwert (SD)	36,3 (4,84)	37,8 (8,68)	36,7 (5,93)
Median (min;max)	36 (30;45)	42 (24;45)	36 (24;45)
Symptome, die zu der Diagnose führten (> 10 %)			
Ataxie	18 (55)	5 (63)	23 (56)
Verhaltensauffälligkeit	5 (15)	4 (50)	9 (22)
Verminderte feinmotorische Fähigkeiten	3 (9)	4 (50)	7 (17)
Abnahme allgemeiner Fähigkeiten	11 (33)	6 (75)	17 (41)
Sprachschwierigkeiten	17 (52)	7 (88)	24 (59)
Muskelschwäche	1 (3)	3 (38)	4 (10)
Myoklonus	8 (24)	3 (38)	11 (27)
Andere motorische Probleme	12 (36)	1 (13)	13 (32)
Krampfanfälle	22 (67)	8 (100)	30 (73)
Verlust des Sehvermögens	4 (12)	3 (38)	7 (17)

Studie 190-901	Hamburg N=33 *n (%)*	Verona N=8 *n (%)*	Insgesamt N=41 *n (%)*
Keine Symptome angegeben	9 (27)	0	9 (22)
Alter (Monate) bei Diagnosestellung			
n	24	8	32
Mittelwert (SD)	59,7 (18,97)	59,9 (9,08)	59,8 (16,90)
Median (min;max)	56 (35;118)	63 (48;72)	57 (35;118)
Geschlecht, n (%)			
Weiblich	11 (33)	2 (25)	13 (32)
Männlich	20 (61)	4 (50)	24 (59)
Unbekannt	2 (6)	2 (25)	4 (10)
Genotyp, n (%)			
c.622C>Tx2	10 (30)	1 (13)	11 (27)
c.509-1G>Cx2	2 (6)	0	2 (5)
c.622C>T und c.509-1G>C	8 (24)	0	8 (20)
c.622C>T und andere	2 (6)	0	2 (5)
c.622C>T und fehlend	0	3 (38)	3 (7)
c.509-1G>C und andere	6 (18)	0	6 (15)
andere x2	4 (12)	2 (25)	6 (15)
fehlend	1 (3)	2 (25)	3 (7)
HML-Score vor bzw. zum Zeitpunkt der Diagnosestellung			
6	0	0	0
5	4 (12)	0	4 (10)
4	7 (21)	3 (38)	10 (24)
3	1 (3)	3 (38)	4 (10)
2	6 (18)	1 (13)	7 (17)
1	1 (3)	0	1 (2)
0	3 (9)	0	3 (7)
fehlend	11 (33)	1 (13)	12 (29)

Abkürzungen: HML: Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg)

Historischer Vergleich: Studie 190-201/190-202 vs. Studie 190-901 (Datenschnitt: 15. Juni 2016)

Im ISE SAP wurde das Analyseverfahren nach 1:1-Paarbildung festgelegt (1:1-Matching der N=23 ITT-Patienten aus Studie 190-201/202 mit historischen Kontrollen aus dem Pool der N=41 bzw. N=42 passenden Kontrollen aus dem DEM-CHILD-Register). Der pU legt im Modul 4 die erste Analyse vom 30. November 2015 vor. Die zugrunde liegenden Daten umfassen den Datenschnitt vom 15. Oktober 2015 der Studie 190-201/190-202 [N=23] und den Datenschnitt vom Februar 2015 der Studie 190-901 [N=41, siehe oben]. Als Begründung verweist der pU im Modul 4 darauf, dass nur diese eine Vergleichbarkeit der Gruppen hinsichtlich prognostisch relevanter Faktoren gewährleisten [2], siehe Tabelle 14. Zusätzlich liegen aktuellere, nicht a priori festgelegte Analysen vom 15. Juni 2016 und vom 01. November 2016 vor. In diese Analysen werden für die Studie 190-201/190-202 der aktuellste Datenschnitt vom 03. Juni 2016 [N=23] und für die Studie 190-901 [N=42] vom August 2016 herangezogen. Somit wurde der Pool der Patienten für die historische Kontrollgruppe um einen Patienten auf N=42 erweitert. Es ist nicht klar, ob dieser eine zusätzliche Patient für das 1:1-Matching der aktuelleren Datenanalysen ausgewählt wurde und ggf. einen anderen Patienten der ersten Analyse des historischen Vergleichs ersetzte. Zudem ergaben sich

für den historischen Vergleich zum Analysezeitpunkt 01. November 2016 N=21 gematchte Paare statt der N=22 bei den früheren Datenschnitten. Im Dossier sind keine aktualisierten Daten hinsichtlich soziodemographischer und krankheitsbezogener Charakteristika für die nach dem 30. November 2015 durchgeführten Analysen angegeben (siehe Tabelle 14). Dennoch wurden für die Nutzenbewertung die Ergebnisse der Analyse vom 15. Juni 2016 extrahiert (190-901: Population möglicher historischer Kontrollen N=42, Datenschnitt August 2016 und 190-201/190-202: ITT N=23, Datenschnitt 03. Juni 2016). Der Vergleich der soziodemographischen und krankheitsbezogenen Charakteristika muss sich allerdings auf die vom pU im Modul 4 dargestellten Angaben der Datenanalyse vom 30. November 2015 (190-901: Population möglicher historischer Kontrollen N=41, Datenschnitt Februar 2015 und 190-201/190-202: ITT N=23, Datenschnitt 15. Oktober 2015) beziehen.

Die Kriterien zur 1:1-Paarbildung („matching“) umfassten einen identischen ML- bzw. HML-Wert und einen Altersunterschied ≤ 12 Monate zu Baseline (300-mg-Baseline in Studie 190-201). Bei Patienten der Kontrollstudie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung. Insgesamt 22 der 23 Patienten der ITT-Population (Studie 190-201) und 22 der 41 „auswertbaren“ Patienten der historischen Kontrollgruppe (Studie 190-901) erfüllten die Kriterien zur 1:1-Paarbildung (siehe Tabelle 14). Bei einem Patienten der 23 ITT-Patienten war eine Paarbildung nicht möglich, da der Altersunterschied zu groß war (21 Monate). Es ist unklar, warum in der historischen Kontrollgruppe (siehe Tabelle 13) keine Patienten einen HML-Score von 6 hatten, im Rahmen des Matchings (190-201/190-202 vs. 190-901) jedoch zwei Personen einen Score von 6 aufwiesen (siehe Tabelle 14).

Zur Analyse vom 15. Juni 2016 lag der mittlere (SD) Altersunterschied zwischen der Cerliponase-alfa-Gruppe und der historischen Kontrollgruppe 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. Die Geschlechterverteilung zwischen den Gruppen war nicht ausgeglichen: Insgesamt 32 % der historischen Kontrollgruppe (190-901) gegenüber 59 % der Studie 190-201/190-202 waren weiblich.

Die Gruppe bestehend aus allen auswertbaren Patienten der Studie 190-901 (historische Kontrolle) hatte zu Baseline einen um etwa einen Punkt schlechteren (niedrigen) ML-Score als die Patienten der Studie 190-201/202. Hinsichtlich des MLV- und Gesamt-Scores zeigte sich ein ähnlicher Trend (siehe Tabelle 14). Ein identischer ML-/HML-Wert war Voraussetzung für das Matching. Es bleibt unklar, warum in der historischen Kontrollgruppe (siehe Tabelle 13) keine Patienten einen HML-Score von 6 aufwiesen, im Rahmen des Matchings (190-201/190-202 vs. 190-901) aber dennoch zwei Personen einen Score von 6 hatten.

Tabelle 14: Charakterisierung der 1:1-Paarbildung für den historischen Vergleich[2)] aus den Studien 190-201/190-202 vs. 190-901

	Studie 190-201/190-202 N=22	**Studie 190-901 N=22**
Alter (Jahre) zu 300 mg Baseline		
Mittelwert (SD)	4,7 (0,91)	4,6 (0,76)
Median (min;max)	4,8 (3,3;7,2)	4,7 (3,3;6,3)
Geschlecht, n (%)		
Weiblich	13 (59)	7 (32)
Männlich	9 (41)	15 (68)
Genotyp, n (%)		
c.622C>T	5 (23)	6 (27)
c.509-1G>C	2 (9)	2 (9)
c.622C>T und c.509-1G>C	2 (9)	5 (23)
c.622C>T und andere	4 (18)	3 (14)
c.622C>T und fehlend	0	1 (5)
c.509-1G>C und andere	3 (14)	2 (9)
andere	6 (27)	3 (14)
ML-(Studien 190-201/202) bzw. HML-Score (Studien 190-901)[1)]		
6		2 (9)
5		2 (9)
4		5 (23)
3		10 (45)
2		2 (9)
1		1 (5)

[1)] Ein identischer ML-/HML-Score war eine Voraussetzung zur Paarbildung. Es ist unklar, warum in der historischen Kontrollgruppe (siehe Tabelle 13) keine Patienten einen HML-Score von 6 hatten, im Rahmen des Matchings (190-201/190-202 vs. 190-901) jedoch zwei Personen einen Score von 6 aufwiesen.

[2)] Methode #1: Identischer ML-/HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline. In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.

Abkürzungen: HML: Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache); ML: Motor/Language (Kombinierte Skala Motorik/Sprache)

2.6.2 Mortalität

Das Gesamtüberleben wurde als UE für die Sicherheitspopulation erfasst. Während der Studie 190-201 und bis zum aktuellen Datenschnitt der Studie 190-201/190-202 (03. Juni 2016) traten keine Todesfälle auf. In der historischen Kontrollstudie 190-901 (DEM-CHILD) wurde das Gesamtüberleben nicht erfasst, so dass ein Vergleich mit der Interventionsstudie nicht möglich ist.

2.6.3 Morbidität

Gemäß SAP (Version 3, vom 25. März 2016) stellte die Responder- und Ereigniszeitanalyse die primäre Auswertung dar. Responder waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 Punkte bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme

von maximal 1 Punkt). Demgegenüber waren Non-Responder (bzw. stabiler Abfall) definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) um zwei oder mehr Punkte (≥ 2) abfiel.

Die hier berichteten Ergebnisse bilden für die Studie 190-201 den Datenschnitt nach 48 Wochen ab (30. November 2015; Studie beendet) und für die Extensionsstudie 190-202 den aktuellsten Datenschnitt vom 3. Juni 2016, wonach 96 % der Patienten über mindestens 72 Wochen mit einer Dosis von 300 mg Cerliponase alfa behandelt wurden. Für den Vergleich der Wirksamkeitsdaten (ML-/HML-Skala) von Studie 190-201/190-202 vs. Studie 190-901 (Integrated Summary of Efficacy, ISE) wird der Datenschnitt vom 15. Juni 2016 herangezogen.

Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016)

Responderanalyse der ML-Skala (primärer Endpunkt)

Die Ergebnisse zum primären Endpunkt sind entsprechend der ITT-Population (N=23) in Tabelle 15 dargestellt. Sowohl zu Woche 48 (Studientag 340 bezogen auf die erste Infusion von 300 mg der Studie 190-201) als auch zu Woche 80 (Studientag 540 bezogen auf die erste Infusion von 300 mg der Studie 190-201) blieb die Responderrate unverändert bei 87 %.

Tabelle 15: Ergebnisse der Responderanalyse der ML-Skala der Studie 190-201 und 190-201/190-202 (ITT-Population, Dosis 300 mg[5])

Studie 190-201/190-202	**Studie 190-201 (Woche 48)[1] N=23**	**Studie 190-201/190-202 (Woche 80)[2] N=23**
Responderanalyse zur ML-Skala *(Netto < 2 bzw. ≥ 2 Punkte oder Score von 0 als Response- bzw. Non-Response-Kriterium)*		
Responder[3], n (% [95%-KI])	20 (87 [66; 97])	20 (87 [66; 97])
Non-Responder[4], n (%)	3 (13)	3 (13)

[1] Datenschnitt: 30. November 2015. Alle Patienten der Studie 190-201 (N=23, ITT) hatten zu diesem Datenschnitt die 48 Wochen der Behandlung mit 300 mg Cerliponase abgeschlossen. Studientag 340 bezogen auf die erste Infusion von 300 mg der Studie 190-201.

[2] Datenschnitt 03. Juni 2016. Der pU gibt im Studienbericht die Woche 80 an, die 6 Beobachtungen enthält, die von Woche 72 übertragen wurden. Die Rationale für die Wahl von 80 Wochen anstelle von 73 Wochen, die für andere Analysen der Studie 190-201/190-202 verwendet wurde, bleibt unklar. Die Responderanalyse bezieht sich auf den Studientag 540 bezogen auf die erste Infusion von 300 mg der Studie 190-201.

[3] Responder waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) und bis Woche 80 (Studientag 540 bezogen auf die erste Infusion von 300 mg) nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme von maximal 1 Punkt) = Ansprechen auf Therapie

[4] Non-Responder (bzw. stabiler Abfall) waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) und bis Woche 80 (Studientag 540 bezogen auf die erste Infusion von 300 mg) im Netto um zwei oder mehr Punkte (≥ 2) abfiel. Ein so genannter stabiler Abfall entsprach einem Abfall um zwei oder mehr Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder um einen Punkt (oder mehr) zugenommen hatte = Fortschreiten der Erkrankung

[5] Baseline ist die letzte Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet).

Abkürzungen: KI: Konfidenzintervall; ML: Motor/Language (Kombinierte Skala Motorik/Sprache)

Sensitivitätsanalysen zu Responderanalysen der ML-Skala (primärer Endpunkt)

A priori wurden im SAP Sensitivitätsanalysen für den primären Endpunkt definiert, deren Ergebnisse nachfolgend in Tabelle 16 dargestellt sind. Die Ergebnisse waren denen der Primäranalyse ähnlich und gleichgerichtet, allerdings numerisch schlechter.

Tabelle 16: Sensitivitätsanalysen zu Responderanalysen der ML-Skala der Studie 190-201 und 190-201/190-202 (Dosis 300 mg[6])

Studie 190-201/190-202	[Analysepopulation] N	Responder, n (%)	[95%-KI]	Non-Responder, n (%)
Studie 190-201 (Datenschnitt: 30. November 2015; zu Woche 48)[3]				
Responderanalyse zur ML-Skala *(Netto < 2 bzw. ≥ 2 Punkte oder Score von 0 als Response- bzw. Non-Response-Kriterium)*	[Sensitivitätsanalyse I][1] N=21	18 (86)	[64%, 97%]	3 (14)
	[Sensitivitätsanalyse II][2] N=18	15 (83)	[59%, 96%]	3 (17)
Responderanalyse zur ML-Skala *(Netto < 1 bzw. ≥ 1 Punkt oder Score von 0 als Response- bzw. Non-Response-Kriterium)*	[ITT-Population] N=23	15 (65)	[43%, 84%]	8 (35)
Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016; zu Woche 80)[4][5]				
Responderanalyse zur ML-Skala zu Woche 80 *(Netto < 2 bzw. ≥ 2 Punkte oder Score von 0 als Response- bzw. Non-Response-Kriterium)*	[Sensitivitätsanalyse I][1][5] N=21	18 (86)	[64%, 97%]	3 (14)
Responderanalyse zur ML-Skala *(Netto < 1 bzw. ≥ 1 Punkt oder Score von 0 als Response- bzw. Non-Response-Kriterium)*	[ITT-Population] N=23	12 (52)	[31%, 73%]	11 (48)

[1] Sensitivitätspopulation I (N=21): Alle Patienten der ITT-Population. Ausgeschlossen wurden Patienten, die zu Baseline einen ML-Score von sechs aufwiesen und im Netto nicht um zwei Punkte (< 2) zu Woche 48 abfielen.
[2] Sensitivitätspopulation II (N=18): Alle Patienten der Sensitivitätsanalyse I und deren ML-Score zu 300-mg-Baseline < 3 Punkte.
[3] Alle Patienten der Studie 190-201 (N=23, ITT) hatten zu diesem Datenschnitt die 48 Wochen der Behandlung abgeschlossen.
[4] Datenschnitt 03. Juni 2016. Bis zu diesem Zeitpunkt waren 96 % der Patienten über mindestens 72 Wochen mit 300 mg Cerliponase alfa behandelt worden. Der pU gibt hier im Studienbericht die Woche 80 an, die 6 Beobachtungen enthält, die von Woche 72 übertragen wurden. Die Rationale für die Wahl von 80 Wochen anstelle von 73 Wochen, die für andere Analysen der Studie 190-201/190-202 verwendet wurde, bleibt unklar. Die Responderanalyse bezieht sich auf den Studientag 540 bezogen auf die erste Infusion von 300 mg der Studie 190-201.
[5] Im Studienbericht 190-202 (Datenschnitt: 03. Juni 2016) finden sich keine weiteren Ergebnisse zu den a priori im SAP definierten Sensitivitätsanalysen.
[6] Baseline ist die letzte Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet).

Abkürzungen: KI: Konfidenzintervall; ML: Motor/Language (Kombinierte Skala Motorik/Sprache)

Ereigniszeitanalyse der ML-Skala

Für den primären Endpunkt wurden zusätzlich Ereigniszeitanalysen nach Kaplan-Meier für die ITT-Population (N=23) durchgeführt. Nachfolgend sind die a priori definierten Analysen *Zeit bis zur stabilen Abnahme (Non-Responder) um ≥ 2 Punkte oder Wert von Null im ML-Score (Non-Responder) in* Abbildung 4 *und Zeit bis zur ersten stabilen Abnahme (Non-Responder) um 1 Punkt im ML-Score in* Abbildung 5 für die Studie 190-201/190-202 (Datenschnitt 03. Juni 2016) dargestellt.

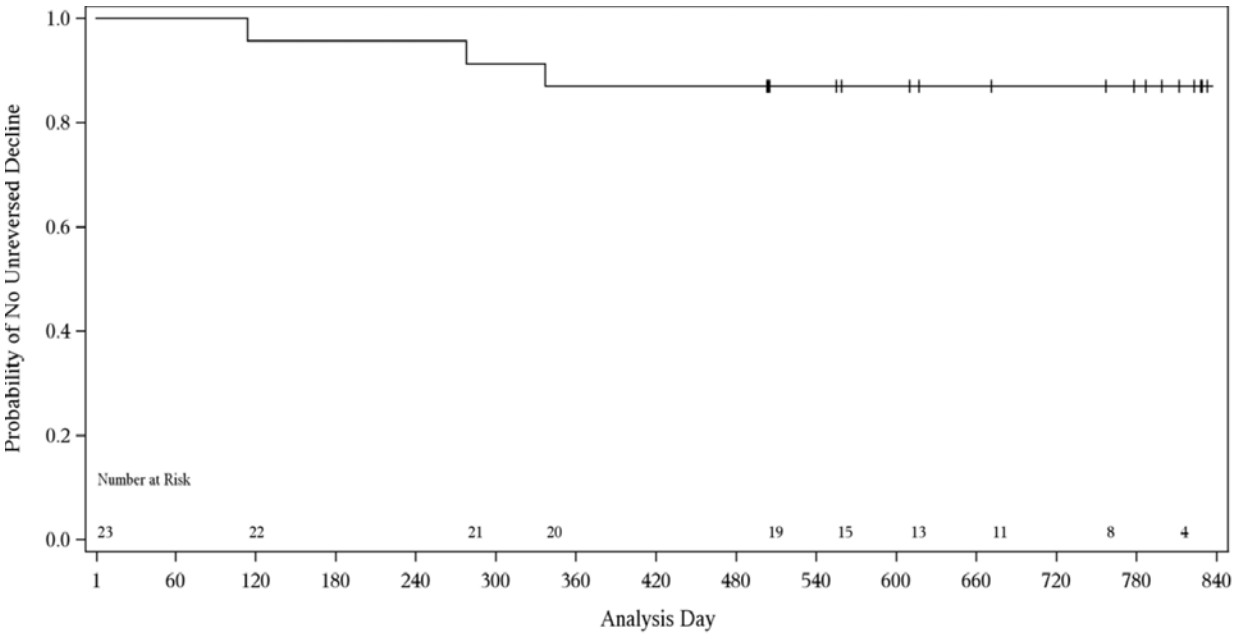


Abbildung 4: Zeit bis zum ersten stabilen Abfall des ML-Scores (≥ 2 Punkte) oder Score von 0 nach Kaplan-Meier der Studie 190-201/190-202 (ITT-Population [N=23], 300-mg-Baseline[1])

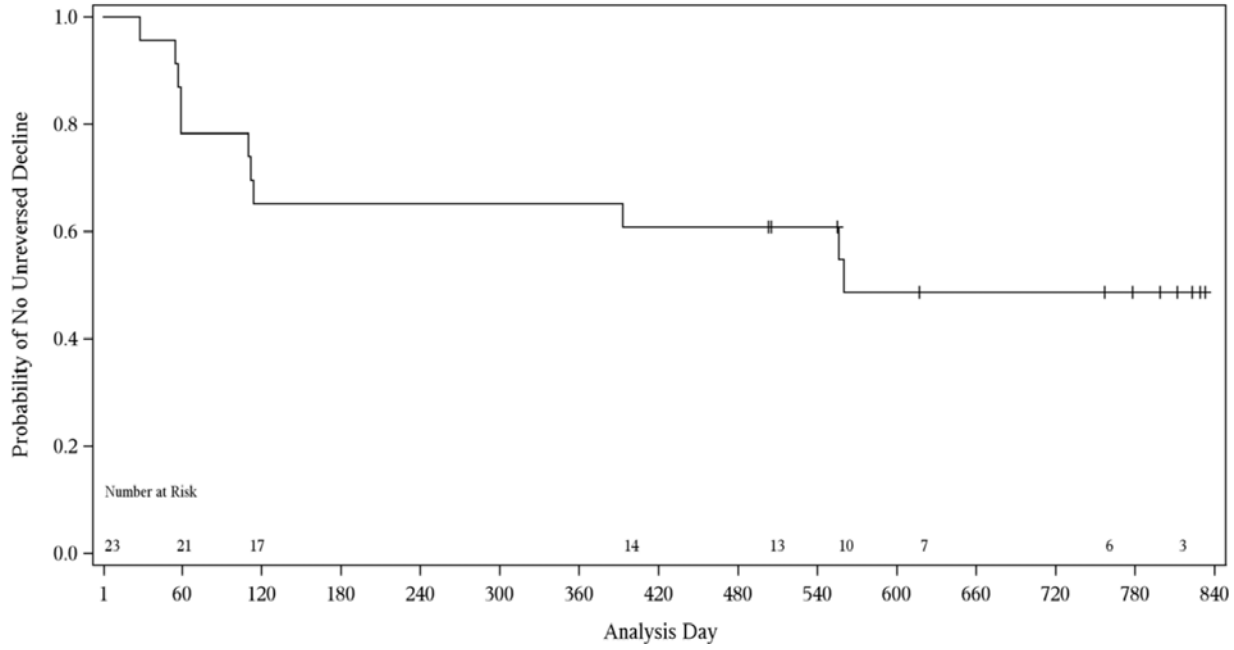


Abbildung 5: Zeit bis zum ersten stabilen Abfall des ML-Scores um 1 Punkt nach Kaplan-Meier für Studie 190-201/190-202 (ITT-Population [N=23], 300-mg-Baseline[2])

[1] Baseline entsprach der letzten Beobachtung vor der ersten 300-mg-Infusion. Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis für Patienten der Dosis-Eskalationsphase nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet.
Tag 1 entspricht dem Tag der ersten 300-mg-Infusion mit Cerliponase alfa [#Studienbericht 190-202 S. 25]

[2] Baseline entsprach der letzten Beobachtung vor der ersten 300-mg-Infusion. Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis für Patienten der Dosis-Eskalationsphase nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet.
Tag 1 entspricht dem Tag der ersten 300-mg-Infusion mit Cerliponase alfa [#Studienbericht 190-202 S. 26]

Slope-Analyse (mittlerer Abfall der ML-Skala, ursprünglicher primärer Endpunkt)

Die Slope-Analyse beschreibt das Außmaß des Abfalls der ML-Skala skaliert für einen Zeitraum von 48 Wochen in der Studie 190-201 und 190-201/190-202. Der mittlere Abfall der ML-Skala skaliert auf 48 Wochen betrug 0,4 Punkte [95%-KI 0,05;0,75] in der Studie 190-201 und 0,32 [95%-KI 0,13;0,52] in der Studie 190-201/190-202 (siehe Tabelle 17).

Tabelle 17: Ergebnisse zur Slope-Analyse der ML-Skala skaliert auf 48 Wochen in der Studie 190-201 und 190-201/190-202 (ITT-Population, Dosis 300 mg[3])

Studie 190-201/190-202	**Studie 190-201[1] N=23**	**Studie 190-201/190-202[2] N=23**
ML-Skala (skaliert auf 48 Wochen)		
MW (SD)	0,40 (0,81)	0,32 (0,46)
Median (min;max)	0 (-0,9;2,0)	0 (-0,4;1,3)
[95%-KI]	[0,05; 0,75]	[0,13; 0,52]

[1] Alle Patienten der Studie 190-201 (N=23, ITT) hatten zu diesem Datenschnitt die 48 Wochen der Behandlung abgeschlossen.
[2] Datenschnitt 03. Juni 2016. Bis zu diesem Zeitpunkt waren 96 % der Patienten über mindestens 72 Wochen mit 300 mg Cerliponase alfa behandelt worden.
[3] Baseline ist die letzte Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet).

Abkürzungen: ML: Motor/Language (Kombinierte Skala Motorik/Sprache)

Sensitivitätsanalysen zur Slopeanalyse (mittlerer Abfall der ML-Skala, ursprünglicher primärer Endpunkt)

A priori wurden im SAP Sensitivitätsanalysen für den primären Endpunkt definiert, deren Ergebnisse nachfolgend in Tabelle 18 dargestellt sind. Die Ergebnisse waren denen der Primäranalyse ähnlich und gleichgerichtet, aber numerisch schlechter.

Tabelle 18: Sensitivitätsanalysen zur Slopeanalyse der ML-Skala skaliert auf 48 Wochen der Studie 190-201 und 190-201/190-202 (Dosis 300 mg[6])

Analysepopulation [N]	**MW (SD)**	**[95%-KI]**	**Median (min;max)**
Studie 190-201(Datenschnitt: 30. November 2015; skaliert auf Woche 48)[3]			
Sensitivitätsanalyse I[1] [N=21]	0,43 (0,84)	[0,05; 0,82]	0 (-0,9;2,0)
Sensitivitätsanalyse II[2] [N=18]	0,51 (0,89)	[0,06; 0,95]	0 (-0,9;2,0)
Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016; skaliert auf Woche 48)[4][5]			
Sensitivitätsanalyse I[1][5] [N=21]	0,35 (0,47)	[0,14; 0,57]	0,4 (-0,4;1,3)

[1] Sensitivitätspopulation I (N=21): Alle Patienten der ITT-Population. Ausgeschlossen wurden Patienten, die zu Baseline einen ML-Score von sechs aufwiesen und im Netto nicht um zwei Punkte (< 2) zu Woche 48 abfielen.
[2] Sensitivitätspopulation II (N=18): Alle Patienten der Sensitivitätsanalyse I mit ML-Score zu 300-mg-Baseline < 3 Punkte.
[3] Alle Patienten der Studie 190-201 (N=23, ITT) hatten zu diesem Datenschnitt die 48 Wochen der Behandlung abgeschlossen.
[4] Datenschnitt 03. Juni 2016. Bis zu diesem Zeitpunkt waren 96 % der Patienten über mindestens 72 Wochen mit 300 mg Cerliponase alfa behandelt worden.
[5] Im Studienbericht 190-202 (Datenschnitt: 03. Juni 2016) finden sich keine weiteren Ergebnisse zu den a priori im SAP definierten Sensitivitätsanalysen.
[6] Baseline ist die letzte Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet).

Abkürzungen: ML: Motor/Language (Kombinierte Skala Motorik/Sprache)

Veränderung der ML-Skala, MLV-Skala und MLVS-Gesamtskala zu Baseline

Entgegen der Einschätzung des pU werden neben der ML-Skala zusätzlich die MLV-Skala (ML-Skala plus Sehvermögen) und die MLVS-Gesamtskala (ML-Skala plus Sehvermögen und Krampfanfälle) als patientenrelevant eingestuft und ergänzend, jedoch nicht als bewertungsrelevant in der Nutzenbewertung dargestellt (siehe Kapitel 2.5.1.). Die Veränderung von 300-mg-Baseline bis Woche 73 und Woche 81 sind für die ITT-Population in Tabelle 19 dargestellt. Bei der ML-Skala und MLV-Skala verschlechterten sich die Werte bis Woche 73, während sie bei der MLVS-Skala unverändert blieben.

Tabelle 19: Ergebnisse zur Veränderung der ML-Skala, MLV-Skala und MLVS-Gesamtskala zu Baseline in der Studie 190-201/190-202 (ITT-Population, Dosis 300 mg[7])

Studie 190-201/190-202	**Studie 190-201/190-202[6] N=23**					
	300-mg-Baseline[1]	**Woche 49**	**Woche 73**	**Veränderung zu Woche 73**	**Woche 81**	**Veränderung zu Woche 81**
ML-Skala (Spanne 0–6 Punkte)[5]						
n	23	23	23	23	17	17
MW (SD)	3,5 (1,2)	3 (1,3)	3 (1,4)	-0,4 (0,9)	2,8 (1,6)	-0,5 (0,87)
Median (min;max)	3 (1;6)	3 (1;6)	3 (1;6)	0 (-2;1)	3 (1;6)	-1 (-2;1)
MLV-Skala (Spanne 0–9 Punkte)[2][4][5]						
n	23	23	23	23	17	17
MW (SD)	6,3 (1,3)	5,7 (1,6)	5,7 (1,8)	-0,6 (1,3)	5,3 (2,1)	-0,9 (1,3)
Median (min;max)	6 (4;9)	6 (3;9)	6 (2;9)	0 (-4;1)	5 (2;9)	-1 (-4;1)
MLVS-Skala (Spanne 0–12 Punkte)[3][4][5]						
n	23	23	23	23	17	17
MW (SD)	8 (1,8)	7,8 (2,1)	8 (2,2)	0,0 (2,2)	7,6 (2,6)	-0,1 (2,6)
Median (min;max)	8 (4;12)	8 (4;12)	8 (4;12)	0 (-5;4)	7 (4;12)	0 (-5;4)

[1] Baseline entsprach der letzten Beobachtung vor der ersten 300-mg-Infusion. Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis für Patienten der Dosis-Eskalationsphase nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet.
[2] Umfasst die ML-Skala plus Sehvermögen (0–9 Punkte).
[3] Umfasst die ML-Skala plus Sehvermögen und Krampfanfälle (0–12 Punkte).
[4] Für jede Einzeldomäne wird der Funktionsstatus auf einer Skala von 3 (normale Funktion) bis 0 (kompletter Verlust der Funktion) erhoben.
[5] Eingesetzte Methode = Schätzung unter Einsatz eines Zeitfensters: Bei den Patienten, die an der Dosiseskalation 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.
[6] Datenschnitt 03. Juni 2016. Bis zu diesem Zeitpunkt waren die ITT-Patienten über mindestens 72 Wochen mit 300 mg Cerliponase alfa behandelt worden. Der pU gibt hier im Studienbericht die Woche 73 und Woche 81 an.
[7] Baseline ist die letzte Beobachtung vor der ersten 300-mg-Infusion (für Patienten der Dosis-Eskalationsphase: Lag eine Beobachtung innerhalb von 8 Wochen vor der ersten 300-mg-Dosis nicht vor, wurde ein Wert innerhalb von 8 Wochen nach der ersten 300-mg-Infusion als Baseline-Wert verwendet).

Abkürzungen: max: Maximum; min: Minimum; ML: Motor/Language (Kombinierte Skala Motorik/Sprache); MW: Mittelwert; SD: Standardabweichung

Historischer Vergleich: Studie 190-201/190-202 vs. 190-901 (Datenschnitt: 15. Juni 2016)

Bereits im Originalprotokoll vom 7. März 2013 [Einschluss erster Patient in die Studie 190-201/190-202: 13. September 2013] war für die Überprüfung der Wirksamkeit ein Vergleich mit einer historischen Kontrolle vorgesehen. Dieser wurde auf Empfehlung der Zulassungsbehörden im ISE SAP (Version 3, 25.04.2016) konkretisiert. In der historischen Vergleichsstudie 190-901 wurde für die Erhebung der Motor- und Sprachfunktionen die ursprünglich entwickelte Skala (HML-Skala) verwendet, während in der Interventionsstudie 190-201/190-202 die adaptierte ML-Skala eingesetzt wurde (nachfolgend als ML-/HML-Skala bezeichnet). Analysiert wurden die Patienten der ITT-Population der Studie 190-201/190-202, denen Patienten der historischen Kontrollgruppe im Verhältnis 1:1 zugeteilt wurden. Voraussetzung für das Matching war, dass der ML-/HML-Wert zu Baseline exakt übereinstimmte und der Altersunterschied so gering wie möglich war (allerdings nicht größer als 12 Monate). Diese Kriterien trafen auf 22 Paare zu, die aus der Studienpopulation 190-201/190-202 (n=23) und der historischen Kontrollgruppe (n=42) gebildet wurden.

Die Definition des primären Endpunkts Response der ISE differiert von der Responderdefinition in der einarmigen Studie 190-201/202: Die Response für den historischen Vergleich der ISE ist definiert anhand des Slope, d. h. anhand der Größe des Abfalls der motorischen und sprachlichen Fähigkeiten, gemessen mittels der Verlaufsdaten der (H)ML-Werte. Der lineare Abfall wurde anhand eines jeweiligen Startwertes, (H)ML-Wert zu Baseline, und eines Endwertes, letzter verfügbarer Messwert > 0, bestimmt, indem diese beiden Punkte zu einer Linie verbunden wurden. Responder waren somit definiert als Patienten, bei denen der Abfall (Slope der Linie), skaliert auf 48 Wochen, kleiner als zwei (< 2) war (festgelegt im finalen SAP ISE, 3. Version vom 25. April 2016).

Die nachfolgend berichteten Ergebnisse beziehen sich auf den Datenschnitt vom 15. Juni 2016. Dabei beziehen sich die Daten der Studie 190-201/190-202 auf den Datenschnitt vom 03. Juni 2016 (die 23 Patienten der ITT-Population waren bis zu diesem Zeitpunkt über mehr als 72 Wochen behandelt worden). Zu Woche 89 liegen für diesen Datenschnitt nur Daten von 11 Patienten der historischen Kontrollgruppe und von 13 Patienten in der Cerliponase-alfa-Gruppe vor (siehe Tabelle 24). Nach Verfügbarkeit werden die Daten aus dem aktuellsten Datenschnitt vom 01. November 2016 mit 21 gematchten Paaren ergänzt (primäre Responderanalyse und Ereigniszeitanalyse). Bis zu diesem Zeitpunkt waren alle Patienten über mindestens 96 Wochen mit Cerliponase alfa in 300 mg Dosierung behandelt worden.

<u>Responderanalyse der ML-/HML-Skala (primärer Endpunkt; 1:1-Paarbildung mit der historischen Kontrolle, Datenschnitt: 15. Juni 2016)</u>

Die Definition des primären Endpunkts Response der ISE differiert von der Responderdefinition in Studie 190-201/202: Im Gegensatz zu Studie 190-201/202 bezieht sich die Response der ISE auf den Slope, d. h. auf die Größe des Abfalls der motorischen und sprachlichen Fähigkeiten, gemessen anhand der Verlaufsdaten der (H)ML-Werte. Der lineare Abfall wurde anhand eines jeweiligen Startwertes, (H)ML-Wert zu Baseline, und eines Endwertes, letzter verfügbarer Messwert > 0, bestimmt, indem diese beiden Punkte zu einer Linie verbunden wurden. Responder waren somit definiert als Patienten, bei denen der Abfall (Slope der Linie), skaliert auf 48 Wochen, kleiner als zwei (< 2) war (festgelegt im finalen SAP ISE, 3. Version vom 25. April 2016).

Die Ergebnisse für die gematchte ITT-Population der Studie 190-201/202 (N=22 Paare) sind in Tabelle 20 dargestellt. Der Anteil der Patienten, bei denen der Abfall skaliert auf 48 Wochen kleiner als zwei war, betrug 100 % (n=22/22) bei den mit Cerliponase alpha behandelten Patienten

verglichen mit 45 % (n=10/22) der Patienten aus der historischen Kontrollgruppe (p=<0,0001). Zum Datenschnitt 01. November 2016 wurden 21 Paare gematcht, von denen alle 21 Patienten (100 %) der Studienpopulation 190-201 unter Studienmedikation ansprachen, während nur 9 Patienten der historischen Kontrolle (43 %) das Responder-Kriterium erfüllten.

Es wurden ebenfalls die Subskalen motorische Fähigkeiten und Sprachfähigkeiten getrennt analysiert, wobei < 1 bzw. ≥ 1 Punkt als Response- bzw. Non-Response-Kriterium definiert waren. Der Unterschied der Ansprechrate betrug auf der motorischen Subskala 45 % (p=0,0005) und hinsichtlich der Sprachdomäne 57 % (p<0,0001).

Tabelle 20: Ergebnisse der Responderanalyse der ML-/HML-Skala der Studie 190-201/190-202 (ITT-Population) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung[3)] (Dosis 300 mg[7)], Datenschnitt: 15. Juni 2016)

Studie 190-201/190-202 vs. 190-901	**Studie 190-201/ 190-202 N=22**	**Studie 190-901 N=22**	**Studie 190-201/190-202 vs. 190-901**	
			RR [95%-KI][6)]	**p-Wert[6)]**
Responderanalyse zur ML-/HML-Skala *(Netto < 2 Punkte/48 Wochen als Response bzw. ≥ 2 Punkte/48 Wochen Non-Response-Kriterium)[1)]*				
< 2 Response-Kriterium[8)]	22 (100)	10 (45)	2,14 [1,37; 3,36]	0,0009
≥ 2 Non-Response-Kriterium	0	12 (55)	0,04 [0,00; 0,64]	0,02
Differenz (190-201) - (190-901)	55 %			-
p-Wert[4)]	< 0,0001			
Motorische Domäne der ML-/HML-Skala *(Netto < 1 Punkt/48 Wochen als Response bzw. ≥ 1 Punkt/48 Wochen als Non-Response-Kriterium)*[2)]				
< 1 Response-Kriterium	22 (100)	12 (55)	1,80 [1,23; 2,63]	0,002
≥ 1 Non-Response-Kriterium	0	10 (45)	0,04 [0,00; 0,64]	0,02
Differenz (190-201) - (190-901)	45 %			-
p-Wert[4)]	0,0005			
Sprachdomäne der ML-/HML-Skala *(Netto < 1 Punkt/48 Wochen als Response bzw. ≥ 1 Punkt/48 Wochen als Non-Response-Kriterium)*[2)]				
n[5)]	21	21	21	21
< 1 Response-Kriterium	21 (100)	9 (43)	2,28 [1,39; 3,75]	0,001
≥ 1 Non-Response-Kriterium	0	12 (57)	0,04 [0,00; 0,63]	0,02
Differenz (190-201) - (190-901)	57 %			-
p-Wert[4)]	< 0,0001			

[1)] Geschätzte Neigung (Slope; Intra-Subjekt) kategorisiert nach < 2 und ≥ 2 Punkte auf der ML/HML-Skala/ 48 Wochen
[2)] Geschätzte Neigung (Slope; Intra-Subjekt) kategorisiert nach < 1 und ≥ 1 Punkt auf der ML/HML-Skala/ 48 Wochen
[3)] Methode #1: Identischer ML-/ HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline. In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.
[4)] Exakter Test nach Fisher
[5)] Ein Paar wurde ausgeschlossen, da der Baseline-Sprachscore bei mindestens einem der Patienten 0 betrug.

[6] Eigene Berechnungen. Korrektur für Zellen mit 0 Ereignissen durch Addition von 0,5.
[7] Baselinewert der Studien 190-201 und 190-202 ist die letzte Beobachtung vor der ersten 300-mg-Infusion von Cerliponase alfa.
[8] Die Definition des primären Endpunkts Response der ISE differiert von der Responderdefinition in der einarmigen Studie 190-201/202: Die Response für den historischen Vergleich der ISE ist definiert anhand des Slope, d. h. anhand der Größe des Abfalls der motorischen und sprachlichen Fähigkeiten, gemessen mittels der Verlaufsdaten der (H)ML-Werte. Der lineare Abfall wurde anhand eines jeweiligen Startwertes, (H)ML-Wert zu Baseline, und eines Endwertes, letzter verfügbarer Messwert > 0, bestimmt, indem diese beiden Punkte zu einer Linie verbunden wurden. Responder waren somit definiert als Patienten, bei denen der Abfall (Slope der Linie), skaliert auf 48 Wochen, kleiner als zwei (< 2) war (festgelegt im finalen SAP ISE, 3. Version vom 25. April 2016).

Abkürzungen: max: Maximum; min: Minimum; ML-/HML-Skala: Motor/Language (Kombinierte Skala Motorik/Sprache)/ Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg)

Sensitivitätsanalysen zur Responderanalyse der ML-/HML-Skala (primärer Endpunkt; 1:1-Paarbildung mit der historischen Kontrolle, Datenschnitt: 15. Juni 2016)

Die Sensitivitätsanalyse I entspricht der A-priori-Definition im SAP und Studienbericht der Studien 190-201/190-202. Die Sensitivitätsanalyse II wurde laut Studienbericht nicht durchgeführt. Zusätzlich wurde eine nicht a priori definierte Analyse aufgenommen (Ausschluss von Patienten, deren Baseline-ML-Score von 6 über den Verlauf der Studie unverändert blieb, und Ausschluss eines gematchten Paares, das zu Baseline einen ML-/HML-Score von 1 aufwies. Die Ergebnisse der Sensitivitätsanalysen sind in Tabelle 21 dargestellt.

Tabelle 21: Sensitivitätsanalysen zur primären Responderanalyse gemessen mittels ML-/HML-Skala: Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung[3)] (Dosis 300 mg[4)], Datenschnitt 15. Juni 2016)

[Analysepopulation] N	Studie 190-201/190-202		Studie 190-901		Studie 190-201/190-202 vs. 190-901
	Responder, n (%)	Non-Responder, n (%)	Responder, n (%)	Non-Responder, n (%)	Differenz p-Wert[3)]
[Sensitivitätsanalyse I][1)] N=20	20 (100)	0	9 (45)	11 (55)	55 % 0,0001
[zusätzliche Sensitivitätsanalyse][2)] N=19	19 (100)	0	8 (42)	11 (58)	58 % 0,0001

[1)] Ausschluss von Patienten mit einem Baseline-ML-Score von 6, der über den Verlauf der Studie unverändert blieb.
[2)] Zusätzlich durchgeführte Sensitivitätsanalyse, die nicht a priori im finalen SAP präspezifiziert wurde. Diese beinhaltet den Ausschluss von Patienten, deren Baseline-ML-Score von 6 über den Verlauf der Studie unverändert blieb, und Ausschluss eines gematchten Paares, das zu Baseline einen ML-/HML-Score von 1 aufwies.
[3)] Methode #1: Identischer ML-/HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline. In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.
[4)] Baselinewert der Studien 190-201 und 190-202 ist die letzte Beobachtung vor der ersten 300-mg-Infusion von Cerliponase alfa. Für den Vergleich mit der historischen Kontrollstudie 190-901 wurden für die Analysen nach 1:1-Paarbildung das Alter (in Monaten zum Zeitpunkt der Untersuchung), zu dem die Paarbildung stattfand, als Baseline festgelegt.

Abkürzungen: ML-/HML-Skala: Motor/Language (Kombinierte Skala Motorik/Sprache)/ Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg)

Darüber hinaus wurden Populationen für eine supportive Analyse definiert. Für den primären Endpunkt finden sich im Datenschnitt vom 15. Juni 2016 ausschließlich Ergebnisse für die

Methode #1 (Simulation) und Methode #2 (1:1-Matching anhand zweitem Auswahlalgorithmus) (nähere Informationen siehe 2.5.4.3). Der über 1.000 Simulationen ermittelte Anteil der Responder, skaliert auf 48 Wochen, betrug 44 % bei den Patienten der historischen Kontrolle (auswertbare Patienten) verglichen mit 100 % bei den Patienten der Cerliponase alfa-Gruppe (ITT). Der Responderanteil bei Methode #2 ist 50 % (11 Patienten) in der historischen Kontrollgruppe und 100 % (22 Patienten) in der Cerliponase-alfa-Gruppe (ITT). Die Ergebnisse der Sensitivitäts- und supportiven Analysen weisen darauf hin, dass die Effekte denen der Hauptanalyse ähnlich sind.

Slope-Analyse (mittlerer Abfall der ML-/HML-Skala; 1:1-Paarbildung mit der historischen Kontrolle, Datenschnitt 15. Juni 2016)

In Tabelle 22 ist die Rate des Abfalls in der ML-/HML-Skala (Progressionsrate), skaliert für einen Zeitraum von 48 Wochen, in der Studie 190-201/202 im Vergleich zur Studie 190-901 dargestellt (1:1-Paarbildung nach Methode #1). Der ML-/HML-Score, skaliert auf 48 Wochen, fiel bei den Patienten der Studie 190-901 im Mittel um 2,0 Punkte ab, verglichen mit 0,34 Punkten bei den Patienten der Studie 190-201/202, was einem Unterschied von 1,66 Punkten (p<0,0001), skaliert auf 48 Wochen, entspricht.

Tabelle 22: Ergebnisse der Slope-Analyse (mittlerer Abfall der ML-/HML-Skala): Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung[4] (Dosis 300 mg[7], Datenschnitt 15. Juni 2016)

[Analysepopulation] (N)	Studie 190-201/ 190-202	Studie 190-901	Studie 190-201/190-202 vs. 190-901	
			Differenz MW (SD) [95%-KI][1]	p-Wert[2]
ML-/HML-Score[3] *(Abfall der ML-/HML-Skala skaliert auf 48 Wochen)*				
n	22	22	1,66 (0,31) [1,02; 2,31]	< 0,0001
MW (SD)	0,34 (0,46)	2,0 (1,39)		
Median (min;max)	0,20 (-0,43;1,33)	2,14 (0;4,98)		
[95%-KI]	[0,13; 0,54]	[1,38; 2,62]		
Motorische Fähigkeiten der ML-/HML-Skala[5] *(Abfall der motorischen Fähigkeiten skaliert auf 48 Wochen)*				
n	22	22	0,74 (0,21) [0,30; 1,18]	0,0020
MW (SD)	0,20 (0,27)	0,93 (0,96)		
Median (min;max)	0 (0;0,67)	0,87 (0;3,73)		
[95%-KI]	[0,08; 0,32]	[0,51; 1,36]		
Sprachvermögen der ML-/HML-Skala[5] *(Abfall des Sprachvermögens skaliert auf 48 Wochen)*				
n[6]	21	21	1,07 (0,21) [0,63;1,51]	< 0,0001
MW (SD)	1,12 (0,93)	0,05 (0,27)		
Median (min;max)	1,24 (0;3,20)	0 (-0,43;0,67)		
[95%-KI]	[0,69; 1,54]	[-0,07; 0,17]		

[1] Zur Berechnung der KIs wurde eine t-Verteilung unter Einsatz der Satterthwaite-Approximation herangezogen.
[2] Zweistichproben-t-Test mit Satterthwaite-Approximation.
[3] 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).

[4] Methode #1: Identischer ML-/HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline. In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.
[5] Die Rate wurde bei jedem Patienten (Intra-Subjekt) aus der Neigung (Slope) der Linie zwischen zwei Motor- bzw. Sprach-Erhebungen (Phase der klinischen Progression) geschätzt.
[6] Ein Paar wurde ausgeschlossen, da der Baseline-Sprachscore bei mindestens einem der Patienten 0 betrug.
[7] Baselinewert der Studien 190-201 und 190-202 ist die letzte Beobachtung vor der ersten 300-mg-Infusion von Cerliponase alfa. Für den Vergleich mit der historischen Kontrollstudie 190-901 wurden für die Analysen nach 1:1-Paarbildung das Alter (in Monaten zum Zeitpunkt der Untersuchung), zu dem die Paarbildung stattfand, als Baseline festgelegt.

Positive Werte entsprechen einer Krankheitsprogression und negative Werte einer Verbesserung des klinischen Zustands.

Abkürzungen: ML-/HML-Skala: Motor/Language (Kombinierte Skala Motorik/Sprache)/ Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg)

Sensitivitätsanalysen zur Slope-Analyse (mittlerer Abfall der ML-/HML-Skala; 1:1-Paarbildung mit der historischen Kontrolle, Datenschnitt: 15. Juni 2016)

Analog den Sensitivitätsanalysen der Responderanalysen wurden ebenfalls nur Ergebnisse der Sensitivitätsanalyse I berichtet, zusätzlich ebenfalls die nicht a priori definierte Analyse (Ausschluss von Patienten, deren Baseline-ML-Score von 6 über den Verlauf der Studie unverändert blieb, und Ausschluss eines gematchten Paares, das zu Baseline einen ML-/HML-Score von 1 aufwies); dargestellt in Tabelle 23. Beide Sensitivitätsanalysen weisen darauf hin, dass die Effekte denen der Primäranalyse ähnlich und gleichgerichtet sind.

Tabelle 23: Sensitivitätsanalysen zur Slope-Analyse (mittlerer Abfall der ML-/HML-Skala): Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung[6] (Dosis 300 mg[5], Datenschnitt 15. Juni 2016)

[Analysepopulation] (N)	Studie 190-201/ 190-202	Studie 190-901	Studie 190-201/190-202 vs. 190-901	
			Differenz MW (SD) [95%-KI][3]	p-Wert[3]
ML-/HML-Score[4] *(Abfall der ML-/HML-Skala skaliert auf 48 Wochen)*				
[Sensitivitätsanalyse I][1] N=20				
n	20	20	1,64 (0,34) [0,94; 2,34]	< 0,0001
MW (SD)	0,37 (0,47)	2,01 (1,44)		
Median (min;max)	0,42 (-0,43;1,33)	2,14 (0;4,98)		
[95%-KI]	[0,15; 0,59]	[1,34; 2,69]		
[zusätzliche Sensitivitätsanalyse][2] N=19 *(Progressionsrate skaliert auf 48 Wochen)*				
n	19	19	1,73 (0,34) [1,02; 2,43]	< 0,0001
MW (SD)	0,39 (0,48)	2,12 (1,40)		
Median (min;max)	0,43 (-0,43;1,33)	2,24 (0;4,98)		
[95%-KI]	[0,16; 0,62]	[1,44; 2,79]		

[1] Ausschluss von Patienten mit einem Baseline-ML-Score von 6, der über den Verlauf der Studie unverändert blieb.
[2] Zusätzlich durchgeführte Sensitivitätsanalyse, die nicht a priori im finalen SAP präspezifiziert wurde. Diese beinhaltet den Ausschluss von Patienten, deren Baseline-ML-Score von 6 über den Verlauf der Studie unverändert blieb, und Ausschluss eines gematchten Paares, das zu Baseline einen ML-/HML-Score von 1 aufwies.
[3] Zweistichproben-t-Test mit Satterthwaite-Approximation. Zur Berechnung der KIs wurde eine t-Verteilung unter Einsatz der Satterthwaite-Approximation (KIs des Unterschieds) herangezogen.
[4] Die Rate wurde bei jedem Patienten (Intra-Subjekt) aus der Neigung (Slope) der Linie zwischen zwei ML-/HML-Erhebungen (Phase der klinischen Progression) geschätzt.
[5] Baselinewert der Studie 190-201 und 190-202 ist die letzte Beobachtung vor der ersten 300-mg-Infusion von Cerliponase alfa.

Für den Vergleich mit der historischen Kontrollstudie 190-901 wurden für die Analysen nach 1:1-Paarbildung das Alter (in Monaten zum Zeitpunkt der Untersuchung), zu dem die Paarbildung stattfand, als Baseline festgelegt.

[6)] Methode #1: Identischer ML-/HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline. In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; In der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.

Positive Werte entsprechen einer Krankheitsprogression und negative Werte einer Verbesserung des klinischen Zustands.

Abkürzungen: ML-/HML-Skala: Motor/Language (Kombinierte Skala Motorik/Sprache)/ Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg); max: Maximum; min: Minimum; MW: Mittelwert; SD: Standardabweichung

Darüber hinaus wurden Populationen für eine supportive Analyse definiert wie oben beim primären Endpunkt Response. Für den Datenschnitt vom 15. Juni 2016 finden sich ausschließlich Ergebnisse für die Methode #2 (nähere Informationen siehe 2.5.4.3). Der Abfall/Slope des ML-/HML-Score, skaliert auf 48 Wochen, war bei den Patienten der historischen Kontrollgruppe 1,96 verglichen mit 0,34 bei den Patienten der Cerliponase-alfa-Gruppe (ITT), was einem Unterschied von 1,62 Punkten/48 Wochen (p<0,0001) entspricht. Die Ergebnisse der Sensitivitäts- und supportiven Analysen weisen darauf hin, dass die Effekte denen der Hauptanalyse ähnlich sind.

Ereigniszeitanalyse der ML-/HML-Skala (Datenschnitt: 01. November 2016)

Ereigniszeitanalysen für Endpunkte Response stabiler Abfall des ML-/HML-Scores

Für die im ISE SAP definierte Ereigniszeitanalyse werden neben den Daten zum Datenschnitt 15. Juni 2016 (N=22 Paare) zusätzlich die aktuellsten Daten vom 01. November 2016 (N=21 Paare) herangezogen. Die Daten wurden mit einem Cox-Modell ausgewertet. Das Hazard Ratio (HR) betrug 0,10 ([95%-KI 0,03; 0,38], p=0,0005) für eine Beobachtungszeit bis 72 Wochen. Somit ergibt sich eine relative Risikoreduktion einer Krankheitsprogression, gemessen anhand des linear geschätzten Abfalls der ML-Skala, von 90 % für Patienten unter Cerliponase alfa im Vergleich zu Patienten der historischen Kontrolle. Die mediane Zeit bis zur Krankheitsprogression 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 Kaplan-Meier-Kurve für die Ereignisse nachhaltiger/stabiler Abfall der ML-/HML-Werte um ≥ 2 Punkte oder auf den Wert von Null (d. h. der Wert fiel beispielsweise nicht um zwei Punkte ab, um anschließend wieder anzusteigen) ist in Abbildung 6 dargestellt. Der Gruppenunterschied mittels eines Cox-Modells war in Woche 96 ebenfalls statistisch signifikant: HR=0,05 ([95%-KI 0,01; 0,17], p<0,0001).

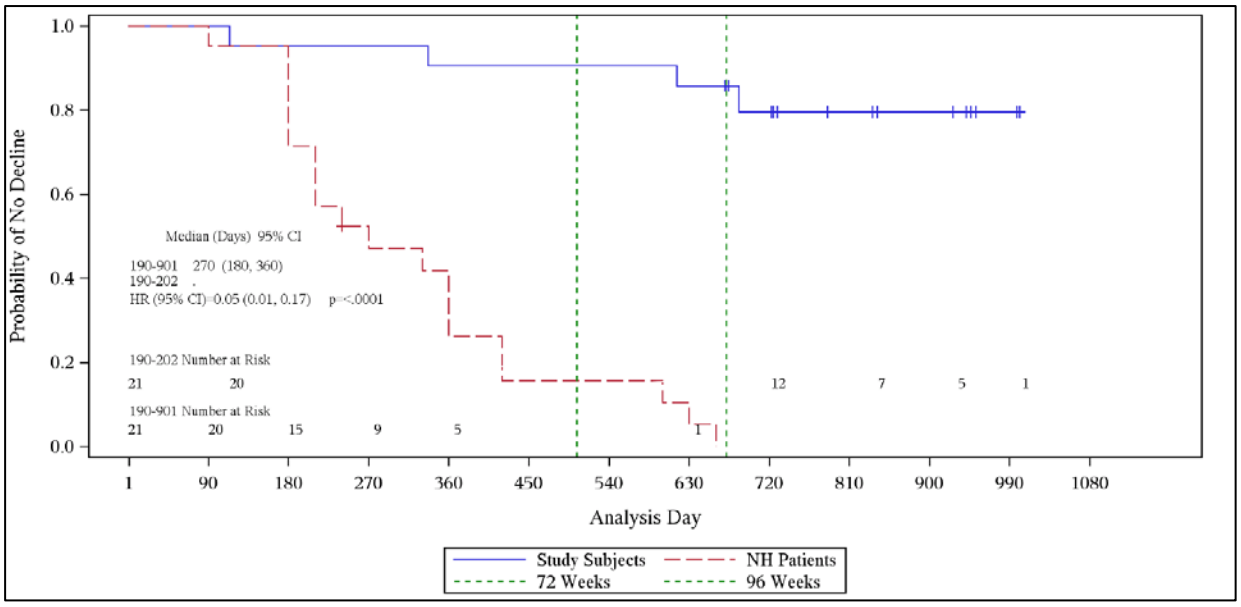


Abbildung 6: Kaplan-Meier-Analyse der Krankheitsprogression gemessen mittels ML-/HML-Skala (stabile Abnahme um ≥ 2 Punkte oder Wert von Null): Cerliponase alfa (190-201/202) vs. unbehandelte Kontrolle (190-901) der ITT-Population mit 1:1-Paarbildung (Alter, ML-/HML-Wert); Datenschnitt 01.11.2016

Zeit, bis der ML-/HML-Score nachhaltig um ≥ 2 Punkte oder auf den Wert von Null abfiel (d. h. der Wert fiel beispielsweise nicht um zwei Punkte mit anschließendem Anstieg ab).
Tag 1 entspricht dem Tag der ersten 300-mg-Infusion mit Cerliponase alfa

<u>Veränderung der ML-/HML-Skala, (H)MLV-Skala und (H)MLVS-Gesamtskala in den Studien 190-201/202 und 190-901 (historische Kontrolle), Datenschnitt: 15. Juni 2016</u>

Zu diesen absoluten Skalenwerten liegen nur deskriptive Daten und Vergleiche vor (siehe Tabelle 24). Der ML-/HML-Score nahm bei den unbehandelten Patienten der historischen Kontrolle im Mittel zu Woche 73 um -2,7 Punkte ab verglichen mit -0,5 Punkten bei den behandelten Patienten. Zu Woche 89 liegen nur Daten von 11 Patienten der historischen Kontrollgruppe und von 13 Patienten in der Cerliponase-alfa-Gruppe vor. Entgegen der Einschätzung des pU werden neben der ML-/HML-Skala zusätzlich die (H)MLV-Skala (ML-Skala plus Sehvermögen) und die (H)MLVS-Gesamtskala (ML-Skala plus Sehvermögen und Krampfanfälle) als patientenrelevant eingestuft, da diese Symptome ebenfalls wesentliche Aspekte der Krankheit betreffen. Aus methodischen Gründen werden diese Ergebnisse in der Nutzenbewertung jedoch nur ergänzend und nicht als bewertungsrelevant dargestellt (siehe Kapitel 2.5.1.). Neben den deskriptiven Angaben zur MLV- und MLVS-Skala lagen jedoch nur Ergebnisse vor, die aus einer nicht-gematchten Analyse stammen, so dass diese nicht berücksichtigt wurden.

Tabelle 24: Ergebnisse zur Veränderung der ML-/HML-Skala zu Baseline: Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung[3)] (Dosis 300 mg, Datenschnitt 15. Juni 2016)

Studie 190-201/190-202 vs. 190-909	Studie 190-201/190-202						Studie 190-901					
	300-mg-Baseline[1)]	Woche 49	Woche 73	Veränderung zu Woche 73	Woche 89	Veränderung zu Woche 89	300-mg-Baseline[1)]	Woche 49	Woche 73	Veränderung zu Woche 73	Woche 89	Veränderung zu Woche 89
ML-/HML-Skala (Veränderung zu Baseline[1)]; 0–6 Punkte, 1:1-Paarbildung Methode #1[2)] [N=23])												
n	22	22	22	22	13	13	22	21	20	20	11	11
MW (SD)	3,4 (1,09)	2,9 (1,19)	2,9 (1,30)	-0,5 (0,91)	2,7 (1,49)	-0,4 (0,76)	3,4 (1,10)	1,2 (1,02)	0,7 (0,66)	-2,7 (0,67)	0,4 (0,54)	-2,7 (0,86)
Median (min;max)	3,0 (1;6)	3,0 (1;6)	3,0 (1;6)	0 (-2;1)	2,1 (1;6)	-0,9 (-1;1)	3 (1;6)	1 (0;3,8)	0,6 (0;2,6)	-2,7 (-3,5;-1,0)	0,2 (0;1,8)	-2,8 (-4,2;-1,0)

[1)] 190-201/202: Baseline = letzter Wert vor der ersten 300-mg-Infusion. 190-901: Baseline = mittlerer Wert für eine Reihe aufeinanderfolgender gleicher Skalenwerte < 6 (oder < 9 HMLV oder Gesamtscore < 12)
[2)] In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.
[3)] Methode #1: Identischer ML-/HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline. In der Studie 190-201/190-202 entsprach Baseline = 300-mg-Baseline; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung.

Abkürzungen: ML-/HML-Skala: Motor/Language (Kombinierte Skala Motorik/Sprache) / Hamburg Motor/Language (Kombinierte Skala Motorik/Sprache, entwickelt in Hamburg); max: Maximum; min: Minimum; MW: Mittelwert; SD: Standardabweichung

2.6.4 Lebensqualität

Zur Erfassung der Lebensqualität wurde der PedsQL™-Fragebogen (PedsQL™-Modul „Parent Report for Toddlers“ und PedsQL™-Modul „Family Impact“ (Eltern-Fremdbericht) herangezogen. Die Ergebnisse, Veränderung zum Studien-Baseline, werden für die einarmige Studie 190-201/190-202 berichtet, da keine Vergleichsdaten der historischen Kohorte 190-901 zur Verfügung stehen. Die Ergebnisse sind in Tabelle 25 dargestellt. Die hier berichteten Ergebnisse bilden den aktuellsten Datenschnitt vom 3. Juni 2016 der Folgestudie 190-202 ab (gesamter Expositionszeitraum einschließlich der Studie 190-201).

PedsQL™-Modul „Parent Report for Toddlers“

Für die Studie 190-201 zeigte sich anhand des PedsQL™-Moduls „Parent Report for Toddlers“ in der ITT-Population (N=23) eine durchschnittliche Verbesserung im Gesamtscore um 2,6 Punkte von Studien-Baseline bis zur letzten Beobachtung in der Studie (Woche 49). Ausschließlich in der Dimension ‚Körperliche Kompetenzen‘ zeigte sich zu Studienende im Vergleich zu Baseline eine Verschlechterung (im Mittel von 6,1 Punkten). In allen Dimensionen zeigte sich zu Woche 49 der Studie 190-202, entsprechend der Woche 98 der Gesamtstudie 190-201/202, im Vergleich zu Baseline der Gesamtstudie eine Verschlechterung. So reduzierte sich der Gesamtscore im Mittel um 14,1 Punkte (siehe Tabelle 25).

Tabelle 25: Ergebnisse zur Lebensqualität – PedsQL™-Modul „Parent Report for Toddlers“ der Studie 190-201/190-202 (ITT) (Studien-Baseline, Dosis 300 mg, Datenschnitt 03. Juni 2016)

Studie 190-201/190-202	Studie 190-201 (N=23)				Studie 190-201/190-202 (N=23)	
	Baseline-Wert[1)]	Woche 25	Woche 49	Veränderung von Baseline[1)] bis Woche 49	Woche 98	Veränderung von Baseline[1)] bis Woche 98
Körperliche Kompetenzen[2)]						
n	23	23	23	23	12	12
MW (SD)	63,5 (22,15)	63 (24)	57,3 (23,6)	-6,1 (21,7)	40,6 (21,9)	-27,9 (22,4)
Median (min;max)	62,5 (25;100)	56,3 (21,9;100)	59,4 (15,6;100)	0 (-43,8;34,4)	15,6 (81,3)	-32,8 (-65,6;18,8)
Emotionale Kompetenzen [2)]						
n	22	23	23	22	12	11
MW (SD)	70,2 (20,7)	82,8 (12,7)	77,8 (16,1)	7,3 (26,2)	67,9 (16,9)	-0,5 (25,5)
Median (min;max)	70 (15;100)	90 (55;100)	80 (45;100)	5 (-35;60)	70 (45;90)	0 (-50;35)
Soziale Kompetenzen[2)]						
n	23	23	23	23	12	12
MW (SD)	49,8 (15,9)	55,7 (15,8)	57,6 (15,3)	7,8 (16,2)	42,5 (16,2)	-5,8 (18,3)
Median (min;max)	50 (25;85)	55 (20;80)	60 (30;90)	10 (-35;40)	45 (15;70)	-2,5 (-40;20)
Schulische Kompetenzen[2)]						
n	22	22	22	22	11	11
MW (SD)	57 (19,6)	72 (17)	63,3 (23,7)	6,3 (29,8)	50 (13,9)	-17 (21,8)
Median (min;max)	58,3 (25;100)	75 (25;91,7)	66,7 (25;100)	0 (-41,7;75)	50 (35,3;75)	-16,7 (-50;12,5)

Studie 190-201/190-202	Studie 190-201 (N=23)				Studie 190-201/190-202 (N=23)	
	Baseline-Wert[1)]	Woche 25	Woche 49	Veränderung von Baseline[1)] bis Woche 49	Woche 98	Veränderung von Baseline[1)] bis Woche 98
Psychosozialer Summenscore[3)]						
n	23	23	23	23	12	12
MW (SD)	59 (12,4)	69,7 (10,1)	67 (13,2)	7,9 (14,9)	53,9 (11,4)	-5,4 (11,1)
Median (min;max)	57,7 (30,8;82,5)	71,2 (38,5;84,6)	65,4 (40,4;94,2)	5,8 (-15,4;36,5)	54,8 (34,6;71,2)	-4,8 (-28,8;9,6)
Gesamtscore[4)]						
n	23	23	23	23	12	12
MW (SD)	60,7 (12,8)	67,2 (12,7)	63,3 (15,2)	2,6 (12,2)	48,7 (13,7)	-14,1 (13,1)
Median (min;max)	59,5 (40,5;81,9)	66,7 (45,2;86,9)	61,9 (39,2;95,2)	2,4 (-17,9;35,7)	47,6 (27,4;73,8)	-16,7 (-32,1;13,1)

[1)] Baseline ist der letzte Wert vor der ersten Infusion (Studien-Baseline).
[2)] 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. Die zu erreichende Punktzahl liegen zwischen 0 und 100. Höhere Punktzahlen deuten auf eine bessere Lebensqualität hin.
[3)] Psychosozialer Summenscore setzt sich aus den Dimensionen Emotionale Kompetenzen, Soziale Kompetenzen und Schulische Kompetenzen zusammen.
[4)] Summe aller Dimensionen: Körperliche Kompetenzen, Emotionale Kompetenzen, Soziale Kompetenzen und Schulische Kompetenzen

Abkürzungen: max: Maximum; min: Minimum; MW: Mittelwert; SD: Standardabweichung

2.6.5 Subgruppenanalysen

Gemäß SAP (Version 3.0 vom 25. März 2016) konnten für die Studie 190-201/190-202 Subgruppen-Merkmale, u. a. häufiger Genotyp (insbesondere c.622C>T- und c.509-1G>C-Mutationen), demographische Merkmale und Studienzentrum oder geographische Region hinsichtlich einer möglichen Effektmodifikation (Wirksamkeit) untersucht werden (siehe Tabelle 1). Im Rahmen der Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016) wurden UEs differenziert für die Merkmale: Alter (≤ 5; > 5), Geschlecht, Land (Deutschland, UK, Italien und USA) und Ethnizität untersucht. Sowohl in der Studie 190-201 als auch in der Extensionsstudie 190-201/202 zeigten sich für die UEs keine wesentlichen Unterschiede in der Inzidenz und Schwere der UEs zwischen den Subgruppen.

Entsprechend ISE SAP konnten verschiedene Intervalle der Behandlungsperiode 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. Im Studienbericht vom 15. Juni 2016 wurden keine der a priori definierten oder darüber hinausgehende Subgruppenanalysen aufgeführt.

2.6.6 Sicherheit

Zur Darstellung von sicherheitsrelevanten Endpunkten wurde die Sicherheitspopulation herangezogen, die alle Patienten umfasst, denen ein ICV-Port implantiert worden war (N=24). 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.

Die mittlere (SD) Dauer der Exposition mit Cerliponase alfa lag bei 47,1 Wochen (10,9) in der 300-mg-Dosierungsperiode während der Studie 190-201 und bei median 95 Wochen für die Studie 190-201/190-202 zum Datenschnitt 03. Juni 2016.

Tabelle 26: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Studie 190-201/190-202

Patienten mit mindestens einem …, n (%)	**Studie 190-201[1)]**	**Studie 190-201/ Studie 190-202[2)]**
UE	24 (100)	24 (100)
UE des NCI-CTCAE		
Grad 1	1 (4)	1 (4)
Grad 2	14 (58)	10 (42)
Grad 3	9 (38)	12 (50)
Grad 4	-	1 (4)
SUE	16 (67)	19 (79)
SUE des NCI-CTCAE		
Grad 1	3 (13)	3 (13)
Grad 2	5 (21)	4 (17)
Grad 3	8 (33)	11 (46)
Grad 4	-	1 (4)

Patienten mit mindestens einem …, n (%)	**Studie 190-201[1)]**	**Studie 190-201/ Studie 190-202[2)]**
UE, das zum Abbruch der Studie führte	0	0
UE, das zum Tod führte	0	0

[1)] Erhebung von unerwünschten Ereignissen (UE) und schwerwiegenden UE (SUE) erfolgte nach schriftlicher Einverständniserklärung und Implantation eines ICV-Ports bis zum Studienende am 30. November 2015. Die Daten umfassen die Gesamtanzahl, unabhängig von der jeweiligen Dosisphase (Dosiseskalations- oder stabile Dosisphase).
[2)] Datenschnitt: 03. Juni 2016; Kumulative Daten aus den Studien 190-201 und 190-202

Abkürzungen: NCI-CTCAE: National Cancer Institute - Common Terminology Criteria for Adverse Events; UE: unerwünschtes Ereignis; SUE: schwerwiegendes UE

Ereignisse mit einer Inzidenz von ≥ 10 %

Unerwünschte Ereignisse mit einer Inzidenz von 10 % oder höher sind nach Systemorganklassen (SOC) und Preferred Term (PT) in Tabelle 27 dargestellt. Folgende SOCs wurden während der Studie 190-201 und dem Datenschnitt der Studie 190-201/190-202 (Datenschnitt: 03. Juni 2016) am häufigsten berichtet: „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 %) und „Psychiatrische Erkrankungen” (50 %/50 %). Dabei stellen folgende UEs (> 40 %) die am häufigsten berichteten PTs dar: Fieber [Pyrexie] (54 %/71 %), Krampfanfall (beide 58 %), Erbrechen (46 %/58 %), Infektion der oberen Atemwege (42 %/50 %), generalisierte tonisch-klonische Anfälle (8 %/46 %), Epilepsie (beide 46 %) und Rhinitis (21 %/42 %).

Tabelle 27: UE mit Inzidenz ≥ 10 % der Studie 190-201/190-202

MedDRA Systemorganklasse, Preferred Term[3)], n (%)	**Studie 190-201**	**Studie 190-201/ Studie 190-202[2)]**
Erkrankungen des Blutes und Lymphsystems[4)]	**3 (13)**	**4 (17)**
Kardiale Erkrankungen[4)]	**4 (17)**	**4 (17)**
Augenerkrankungen[4)]	**3 (13)**	**4 (17)**
Erkrankungen des Gastrointestinaltraktes	**19 (79)**	**19 (79)**
Erbrechen	11 (46)	14 (58)
Obstipation	7 (29)	8 (33)
Diarrhö	4 (17)	6 (25)
Dysphagie	4 (17)	5 (21)
Zahnkaries	2 (8)	3 (13)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	**16 (67)**	**19 (79)**
Pyrexie	13 (54)	17 (71)
Gangstörungen	7 (29)	7 (29)
Entwicklungsverzögerung	3 (13)	4 (17)
Erkrankung des Immunsystems	**8 (33)**	**10 (42)**
Überempfindlichkeitsreaktionen	8 (33)	9 (38)

MedDRA Systemorganklasse, Preferred Term[3], n (%)	**Studie 190-201**	**Studie 190-201/ Studie 190-202[2]**
Infektionen und parasitäre Erkrankungen	**23 (96)**	**24 (100)**
Infektion der oberen Atemwege	10 (42)	12 (50)
Rhinitis	5 (21)	10 (42)
Nasopharyngitis	7 (29)	9 (38)
Virale Infektion	5 (21)	8 (33)
Gastroenteritis	5 (21)	7 (29)
Pharyngitis	6 (25)	6 (25)
Tonsillitis	2 (8)	5 (21)
Konjunktivitis	2 (8)	4 (17)
Infektion der Atemwege	2 (8)	4 (17)
Virale Infektion der oberen Atemwege	n.b.	6 (25)
Rhinovirus-Infektion	2 (8)	3 (13)
Harnwegsinfektion	2 (8)	3 (13)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	**17 (71)**	**19 (79)**
Sturz	6 (25)	7 (29)
Kopfverletzung	2 (8)	3 (13)
Untersuchungen[4]	**3 (13)**	**6 (25)**
Stoffwechsel- und Ernährungsstörungen	**5 (21)**	**9 (38)**
Verminderter Appetit	2 (8)	3 (13)
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	**1 (4)**	**5 (21)**
Muskelspasmen	n.b.	3 (13)
Erkrankungen des Nervensystems	**24 (100)**	**24 (100)**
Krampfanfall	14 (58)	14 (58)
Epilepsie	11 (46)	11 (46)
Generalisierte tonisch-klonische Anfälle	2 (8)	11 (46)
Myoklonus	7 (29)	7 (29)
Petit-mal-Epilepsie	2 (8)	7 (29)
Tremor	4 (17)	7 (29)
Dystonie	4 (17)	5 (21)
Babinski-Reflex	4 (17)	5 (21)
Kopfschmerzen	3 (13)	4 (17)
Athetose	n.b.	3 (13)
Pleozytose	3 (13)	3 (13)

MedDRA Systemorganklasse, Preferred Term[3], n (%)	**Studie 190-201**	**Studie 190-201/ Studie 190-202[2]**
Produktprobleme	**n.b.**	**7 (29)**
Problem mit einer Nadel	n.b.	5 (21)[1]
Psychiatrische Erkrankungen	**12 (50)**	**12 (50)**
Insomnie	4 (17)	4 (17)
Reizbarkeit	4 (17)	4 (17)
Anomales Verhalten	3 (13)	4 (17)
Schlafstörungen	3 (13)	3 (13)
Erkrankungen der Atemwege, des Brustraums und Mediastinums	**6 (25)**	**10 (42)**
Husten	5 (21)	7 (29)
Atemwegserkrankung	n.b.	3 (13)
Erkrankungen der Haut und des Unterhautzellgewebes[4]	**5 (21)**	**6 (25)**
Gefäßerkrankungen[4]	**2 (8)**	**4 (17)**

[1] 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".
[2] Aktueller Datenschnitt der Studie 190-202: 3. Juni 2016. Es handelt sich hierbei um kumulative Daten der Studie 190-201 und 190-202.
[3] Bei Patienten, bei denen mehr als ein UE innerhalb einer MedDRA (SOC oder PT) mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt.
[4] Kein Preferred Term wurde mit ≥ 10 % der Patienten beobachtet.

Abkürzungen: n.b.: nicht berichtet; UE: unerwünschtes Ereignis

UE des CTCAE-Grades 3 oder höher

Unerwünschte Ereignisse des NCI-CTCAE-Grades 3 und höher sind nach Systemorganklassen (SOC) und Preferred Term (PT) in Tabelle 28 dargestellt. Während der Studie 190-201 hatten neun Patienten ein Grad-3-UE, aber kein Grad-4-UE. Bis zum Datenschnitt im Juni 2016 hatten 12 Patienten ein Grad-3-UE und ein Patient ein Grad-4-UE entwickelt (N=13 [54 %]). Ein Patient (4 %) entwickelte einen Status epilepticus (Grad 4), dieser war aus Sicht des Prüfarztes allerdings nicht behandlungsbedingt.

Tabelle 28: UE des NCI-CTCAE-Grades 3 und höher ≥ 2 %

MedDRA Systemorganklasse, Preferred Term[3], n (%)	**Studie 190-201[1]**	**Studie 190-201/ Studie 190-202[2]**
Erkrankungen des Blutes und Lymphsystems	**1 (4)**	**4 (17)**
Thrombozytopenie	1 (4)	2 (8)
Anämie	1 (4)	1 (4)
Erkrankungen des Gastrointestinaltraktes	n.b.	**1 (4)**
Dysphagie	n.b.	1 (4)
Erkrankung des Immunsystems	**3 (13)**	**3 (13)**
Überempfindlichkeitsreaktionen	3 (13)	3 (13)

MedDRA Systemorganklasse, Preferred Term[3], n (%)	**Studie 190-201[1]**	**Studie 190-201/ Studie 190-202[2]**
Infektionen und parasitäre Erkrankungen	**3 (13)**	**6 (25)**
Clostridium-difficile-Kolitis	1 (4)	1 (4)
Gerätebezogene Infektion	n.b.	1 (4)
Pharyngitis	1 (4)	1 (4)
Propionibacterium-Infektion	1 (4)	1 (4)
Hautinfektion	n.b.	**1 (4)**
Infektion der oberen Atemwege	n.b.	2 (8)
Pneumonie	1 (4)	1 (4)
Stoffwechsel- und Ernährungsstörungen	**1 (4)**	**2 (8)**
Acidose	n.b.	1 (4)
Verminderter Appetit	1 (4)	1 (4)
Erkrankungen des Nervensystems	**2 (8)**	**4 (17)**
Krampfanfälle	n.b.	1 (4)
intrakranielle Hämorrhagie	1 (4)	1 (4)
Motorische Dysfunktion	1 (4)	1 (4)
Status epilepticus	n.b.	1 (4)
Erkrankung der Geschlechtsorgane und der Brustdrüse	**1 (4)**	**1 (4)**
Vaginalausfluss	**1 (4)**	**1 (4)**
Erkrankungen der Atemwege, des Brustraums und Mediastinums	**2 (8)**	**3 (13)**
Aspiration	**1 (4)**	1 (4)
Schlafapnoe-Syndrom	**1 (4)**	1 (4)
Mandelhypertrophie	n.b.	1 (4)
Chirurgische und medizinische Verfahren	**n.b.**	1 (4)
Zahnreparatur	n.b.	1 (4)

[1] Beziehen sich auf mindestens ein mit der Behandlung in Zusammenhang stehendes UE (N=23).
[2] Aktueller Datenschnitt der Studie 190-202: 03. Juni 2016. Es handelt sich hierbei um kumulative Daten der Studie 190-201 und 190-202.
[3] Bei Patienten, bei denen mehr als ein UE innerhalb einer MedDRA (SOC oder PT) mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt.

Abkürzungen: n.b.: nicht berichtet; UE: unerwünschtes Ereignis

Schwerwiegende UE (SUE)

Schwerwiegende UEs mit einer Inzidenz von 5 % oder höher sind nach Systemorganklassen (SOC) und Preferred Term (PT) in Tabelle 29 dargestellt. Während der Studie 190-201 traten bei 16 Patienten (67 %) insgesamt 32 SUEs auf. Bis zum Datenschnitt im Juni 2016 berichteten 19 der 24 Patienten (79 %) von insgesamt 51 SUEs.

Tabelle 29: SUE mit Inzidenz ≥ 5 %

MedDRA Systemorganklasse, Preferred Term[3)], n (%)	**Studie 190-201**	**Studie 190-201/ Studie 190-202[2)]**
Erkrankungen des Gastrointestinaltraktes[4)]	**1 (4)**	**19 (79)**
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort[1)]	**2 (8)**	**2 (8)**
Pyrexie	2 (8)	2 (8)
Erkrankung des Immunsystems	**6 (25)**	**7 (29)**
Überempfindlichkeitsreaktionen	6 (25)	7 (29)
Infektionen und parasitäre Erkrankungen	**7 (29)**	**13 (54)**
Bakterielle Pharyngitis	2 (8)	2 (8)
Gastroenteritis	1 (4)	3 (13)
Infektion der oberen Atemwege	n.b.	4 (17)
Verletzungen, Vergiftung und durch Eingriffe bedingte Komplikationen[4)]	**2 (8)**	**2 (8)**
Erkrankungen des Nervensystems	**4 (17)**	**6 (25)**
Epilepsie	2 (8)	2 (8)
Erkrankungen der Atemwege, des Brustraums und Mediastinums[4)]	**1 (4)**	**2 (8)**

[1)] 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“.
[2)] Aktueller Datenschnitt der Studie 190-202: 3. Juni 2016. Es handelt sich hierbei um kumulative Daten der Studie 190-201 und 190-202.
[3)] Bei Patienten, bei denen mehr als ein UE innerhalb einer MedDRA (SOC oder PT) mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt.
[4)] Kein Preferred Term wurde mit ≥ 5 % der Patienten beobachtet.

Abkürzungen: n.b.: nicht berichtet; UE: unerwünschtes Ereignis

Vom pU beschriebene UE von besonderem Interesse

UEs von besonderem Interesse sind nach Systemorganklassen (SOC) und Preferred Term (PT) in Tabelle 30 dargestellt. Die Daten beziehen sich auf den Datenschnitt vom 03. Juni 2016.

(1) Inzidenz von Überempfindlichkeits- und anaphylaktischen Reaktionen

Alle berichteten Überempfindlichkeitsreaktionen wurden durch die breite MedDRA-SMQ „Überempfindlichkeit“ identifiziert und kein UE durch die SMQ „Anaphylaktische Reaktionen“. 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 %). Zwischen dem Abschlussbericht der Studie 190-201 und dem Datenschnitt im Juni 2016 wurden insgesamt 12 weitere Überempfindlichkeitsreaktionen berichtet, allerdings blieb die Zahl der Patienten mit Überempfindlichkeitsreaktionen unverändert (n=14/24, 58 %).

(2) Inzidenz von UEs in zeitlichem Zusammenhang mit den Infusionen („temporally-related events", TRE)

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 (N=24).

Die meisten TREs wurden innerhalb der SOCs „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 [Pyrexie] (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.

(3) UEs in Zusammenhang mit dem ICV-Port

UEs im Zusammenhang mit dem ICV-Port waren nicht a priori als UEs von besonderem Interesse definiert, sondern erst im Amendement 3 (26. Februar 2016) der Studie 190-202 als UEs besonderen Interesses definiert. Insgesamt 11 Patienten (46 %) berichteten von insgesamt 30 UEs, die im Zusammenhang mit dem ICV-Port standen. 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 %) und Apparatur-Leck (5 UEs bei 1 Patient, 4 %) wurden während der Studie 190-201/202 mehr als einmal berichtet. 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.

(4) 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 wurden durch eine breite MedDRA-Abfrage „Konvulsionen" (SMQ) identifiziert.

Zum Datenschnitt 3. Juni 2016 traten keine Fälle von Hydrozephalus und Meningitis auf. Alle Patienten entwickelten mindestens ein UE innerhalb der SOC „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.

Im Rahmen der breiten MedDRA-Abfrage „Konvulsionen" (SMQ) wurden bis zum Datenschnitt 03. Juni 2016 insgesamt 362 UEs bei 23/24 Patienten (96 %) identifiziert. 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 %]). Sechs Patienten (25 %) entwickelten in der Studie 190-201 insgesamt 7 Konvulsionen, die nach der Implantation des ICV-Ports, aber vor der ersten Dosis Cerliponase alfa auftraten.

Tabelle 30: UEs von besonderem Interesse ≥ 10 %[6)]

MedDRA Systemorganklasse, Preferred Term[5)], n (%)	**Studie 190-201**	**Studie 190-201/ Studie 190-202[4)]**
(1) Inzidenz von Überempfindlichkeits- und anaphylaktischen Reaktionen[1)]		
Patienten mit mindestens einem berichteten Überempfindlichkeitsreaktion-UE	14 (58)	14 (58)
Erkrankung des Immunsystems	8 (33)	10 (42)
Überempfindlichkeitsreaktionen	8 (33)	9 (38)
Infektionen und parasitäre Erkrankungen	2 (8)	4 (17)
Konjunktivitis	2 (8)	4 (17)
Erkrankungen der Haut und des Unterhautzellgewebes	5 (21)	5 (21)
(2) Inzidenz von UEs in zeitlichem Zusammenhang mit den Infusionen („temporally-related events", TRE)[2)]		
Patienten mit mindestens einem UE, das in zeitlichem Zusammenhang mit der Infusion auftrat	24 (100)	24 (100)
Erkrankungen des Blutes und Lymphsystems	3 (13)	3 (13)
Kardiale Erkrankungen	2 (8)	3 (13)
Erkrankungen des Gastrointestinaltraktes	10 (42)	12 (50)
Obstipation	3 (13)	4 (17)
Diarrhö	1 (4)	3 (13)
Erbrechen	5 (21)	7 (29)
Allgemeine Erkrankungen und Erkrankungen am Verabreichungsort	11 (46)	11 (46)
Pyrexie	11 (46)	11 (46)
Gangstörungen	4 (17)	4 (17)
Erkrankung des Immunsystems	8 (33)	9 (38)
Überempfindlichkeitsreaktionen	8 (33)	9 (38)
Infektionen und parasitäre Erkrankungen	8 (33)	15 (63)
Gastroenteritis	n.b.	3 (13)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	4 (17)	5 (21)
Sturz	1 (4)	3 (13)
Untersuchungen	3 (13)	4 (17)
Stoffwechsel- und Ernährungsstörungen	3 (13)	6 (25)
Erkrankungen des Nervensystems	16 (67)	19 (79)
Generalisierte tonisch-klonische Anfälle	n.b.	3 (13)
Kopfschmerzen	3 (13)	4 (17)
Myoklonus	4 (17)	4 (17)
Krampfanfälle	8 (33)	8 (33)
Tremor	3 (13)	6 (25)

MedDRA Systemorganklasse, Preferred Term[5], n (%)	**Studie 190-201**	**Studie 190-201/ Studie 190-202[4]**
Produktprobleme	n.b.	5 (21)
Problem mit einer Nadel	n.b.	5 (21)
Psychiatrische Erkrankungen	7 (29)	7 (29)
(3) Inzidenz von UEs, die im Zusammenhang mit dem ICV-Zugang stehen[7]		
Patienten mit mindestens einem UE, das im Zusammenhang mit dem ICV-Port steht	7 (29)	11 (46)
Allgemeine Erkrankungen und Erkrankungen am Verabreichungsort	3 (13)	2 (8)
Erkrankungen des Nervensystems	3 (13)	3 (13)
Pleozytose	3 (13)	3 (13)
Produktprobleme	n.b.	7 (29)
Problem mit einer Nadel	n.b.	4 (17)
(4) Inzidenz von UEs innerhalb der SOC „Erkrankungen des Nervensystems“ und Konvulsionen[3]		
Patienten mit mindestens einem UE, das innerhalb der SMQ Konvulsionen aufgetreten war	23 (96)	23 (96)
Erkrankungen des Nervensystems (SMQ „Konvulsionen“)	23 (96)	23 (96)
Epilepsie	11 (46)	11 (46)
Generalisierte tonisch-klonische Anfälle	2 (8)	11 (46)
Petit-mal-Epilepsie	2 (8)	7 (29)
Krampfanfall	14 (58)	14 (58)

[1] 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.
[2] 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.
[3] Status epilepticus, Hydrozephalus und Meningitis waren a priori als UEs von besonderem Interesse definiert. Zusätzliche UEs wurden durch eine breite MedDRA-Abfrage „Konvulsionen“ (SMQ) identifiziert.
[4] Aktueller Datenschnitt der Studie 190-202: 3. Juni 2016. Es handelt sich hierbei um kumulative Daten der Studie 190-201 und 190-202.
[5] Bei Patienten, bei denen mehr als ein UE innerhalb einer MedDRA (SOC oder PT) mehrmals auftrat, wurde jeweils das schwerste Ereignis berücksichtigt.
[6] siehe Tabelle 27: UE mit Inzidenz ≥ 10 %
[7] Nachträglich mit Amendement 3 als UE von besonderem Interesse definiert. Die Prüfärzte bewerteten den Zusammenhang mit dem Medizinprodukt (ICV-Zugang).

Abkürzungen: n.b.: nicht berichtet; SMQ: Standardised MedDRA Query (Standardisierte MedDRA-Abfragen); SOC: Systemorganklassen; UE: unerwünschtes Ereignis

<u>(5) UEs von besonderem Interesse: Unerwartete und rasche Krankheitsprogression (gemessen als Abfall auf der ML-Skala), die nicht anders erklärbar ist</u>

In der Studie 190-201 fiel bei keinem der Patienten der ML-Score unerwartet rasch und aus nicht anders erklärbaren Gründen ab. Der Zwischenbericht vom 15.10.2015 bestätigt dies. Im aktuellen Datenschnitt (3. Juni 2016) wurde das UE nicht berichtet.

3 Methodische Anmerkungen und Bewertung der Unterlagen

3.1 Zulassungsstatus und Zulassungspopulation von Cerliponase alfa

Gemäß Zulassung ist Cerliponase alfa (Brineura®) indiziert zur Behandlung der neuronalen Ceroid-Lipofuszinose Typ 2 (CLN2-Erkrankung). Es ist zugelassen als Arzneimittel zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. Der Wirkstoff von Cerliponase alfa wird direkt in das Gehirn infundiert. Die Zulassung durch die Europäische Arzneimittelagentur (EMA) besteht seit dem 30. Mai 2017.

Die vorliegende Nutzenbewertung zu Cerliponase alfa basiert auf der Zulassungsstudie 190-201, die sich in eine Dosis-Eskalationsphase und eine stabile Dosisphase gliedert und bereits abgeschlossen ist, sowie der sich daran anschließenden Extensionsstudie 190-202 mit einer Weiterführung der stabilen Dosis von 300 mg, die noch bis Dezember 2020 andauert. Zusätzlich wurde a priori im Originalprotokoll und SAP der Studie 190-201/190-202 festgelegt, dass für die primäre Auswertung ein Vergleich mit einer historischen Kontrollgruppe vorgesehen war. Diese wurde auf Empfehlung der Zulassungsbehörden im Statistischen Analyseplan der „Integrated Summary of Efficacy“ (SAP ISE) spezifiziert und die Ergebnisse in einem vom pU zusätzlichen eingereichten Studienbericht (ISE-Analyse) berichtet.

Trotz Zulassung ohne Altersbeschränkung wurden in die Studie 190-201 bzw. 190-202 ausschließlich Patienten mit einem Alter von drei bis acht Jahren eingeschlossen (Einschlusskriterien zwischen 3 und 15 Jahren). Weitere Einschlusskriterien waren eine milde bis moderate CLN2-Erkrankung unter der Annahme eines Scores von 3–6 auf der ML-Skala sowie einer bestätigten Diagnose anhand einer CLN2-Genanalyse und TPP1-Enzymaktivität. Ausgeschlossen waren Patienten mit einer schweren Verlaufsform und den oben genannten sehr jungen und älteren Altersgruppen, sodass eine Aussage zu diesen Teilpopulationen auf Basis der vorgelegten Daten nicht möglich ist [18]. Vor diesem Hintergrund und fehlender Langzeitdaten wurde Cerliponase alfa unter „außergewöhnlichen Umständen“ (exceptional circumstances) zugelassen. Die EMA und der pU argumentieren, dass es aufgrund der Seltenheit der Krankheit nicht möglich war, vollständige Informationen über Cerliponase alfa zu erlangen [17]. Die Zulassung unter „außergewöhnlichen Umständen“ ist mit der Auflage weiterer Studienergebnisse verknüpft: Eine nicht-interventionelle post-authorisation Sicherheitsstudie (PASS; Studiennummer 190-501) zur Beurteilung der Langzeitsicherheit einschließlich dem Auftreten von Überempfindlichkeitsreaktionen und Anaphylaxien auf Basis einer adäquaten Datenquelle, die sich aus einer Registrierung von Patienten mit CLN2 ergibt (Registerstudie). Ergebnisse sollen für eine jährliche Neubewertung vorgelegt werden. Desweiteren eine nicht-interventionelle post-authorisation Wirksamkeitsstudie (PAES; Studiennummer 190-203) zur Beurteilung der Krankheitsprogression anhand der motorischen und sprachlichen Skala sowie der Sicherheit und Verträglichkeit. In diese Studie sollen mindestens fünf Patienten unter 2 Jahren eingeschlossen werden, Ergebnisse sind im Jahr 2020 zu erwarten [18].

3.2 Design und Methodik der Studie 190-201/190-202

Bei den Zulassungsstudien handelt es sich um unkontrollierte Studien.

Bei der beendeten Studie 190-201 (Studienende 30. November 2015) handelt es sich um eine ‚first-in-human Phase 1/2-Studie' – eine multizentrische, offene, einarmige Interventionsstudie, die sich in zwei Phasen gliederte: Eine Dosis-Eskalationsphase (d. h. Patienten erhielten initial Dosen von 30, 100 bzw. 300 mg Cerliponase alfa/ alle 2 Wochen über jeweils mindestens vier Wochen) und eine nachfolgende stabile Dosierungsphase (für mindestens 48 Wochen 300 mg Cerliponase alfa/ 2 Wochen). In der sich anschließenden und noch andauernden Extensionsstudie 190-202, einer multizentrischen, offenen, einarmigen Verlängerungsstudie der Studie 190-201, erhalten die Patienten weiterhin die stabile Dosis von 300 mg Cerliponase alfa/ alle 2 Wochen für geplant bis zu 239 Wochen (geplantes Studienende: 15. Dezember 2020).

Die Studie 190-201 hatte zum Ziel, die Sicherheit, Akzeptanz, Pharmakokinetik und Wirksamkeit von Cerliponase alfa bei der spät-infantilen CLN2-Erkrankung zu untersuchen, während die Studie 190-202 die langfristige Wirksamkeit und Sicherheit untersucht. Der primäre Endpunkt war der Anteil von Patienten, der auf die Therapie angesprochen hatte (Responder; ITT-Population). Responder waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme von maximal 1 Punkt). Demgegenüber waren Non-Responder (bzw. stabiler Abfall) definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) um zwei oder mehr Punkte (≥ 2) abfiel. Ein so genannter stabiler Abfall entsprach einem Abfall um zwei oder mehr Punkte, der bis zum letzten Untersuchungszeitpunkt nicht wieder um einen Punkt (oder mehr) zugenommen hatte.

Ein zentrales Einschlusskriterium der Studie 190-201 war die leichte bis mittelschwere CLN2-Erkrankung. Die Erkrankungsschwere wurde mittels adaptierter ML-Skala der ursprünglichen Hamburg-ML-Skala erhoben. Beide ML-Skalen (Spanne 0–6 Punkte) umfassen die Kombination von zwei Domänen: motorische Fähigkeiten und Sprachfähigkeiten[2]. In die Studie eingeschlossen wurden nur Patienten, die einen Score von 3–6 auf der ML-Skala und einen ML-Score von mindestens 1 in jeder der beiden Domänen aufzeigten. Die adaptierte ML-Skala ist gleichzeitig das Erhebungsinstrument für die Endpunkte zur Krankheitsprogression. Die Hamburg-Skala besteht im Original jedoch aus zwei weiteren Domänen: Sehvermögen und Krampfanfälle. Warum die beiden letzteren Domänen von Seiten des pU als nicht patientenrelevant eingestuft wurden, ist in den Studienunterlagen nicht nachhaltig und nachvollziehbar begründet. Insbesondere, da das Sehvermögen als auch Krampfanfälle maßgebliche Kennzeichen dieses Erkrankungsbildes darstellen [23] und insbesondere Krampfanfälle bei der Diagnosestellung in der historischen Kohortenstudie als maßgeblich galten (30 von 40 Patienten [73 %]). Eine Rationale für den Ausschluss von Patienten unter drei Jahren ist nicht nachvollziehbar begründet, insbesondere da die Erkrankung i. d. R. zwischen dem 2. und 4. Lebensjahr auftritt. Dies merkt die EMA ebenfalls

[2] Für jede Einzeldomäne wird der Funktionsstatus auf einer Skala von 3 (normale Funktion) bis 0 (kompletter Verlust der Funktion) erhoben.

an, die folglich eine post-authorisation Wirksamkeitsstudie (190-203) fordert, in die auch mindestens fünf Patienten unter 2 Jahren einzuschließen sind (siehe Kapitel 3.1.).

In die Extensionsstudie 190-202 wurden alle Patienten eingeschlossen, die die 48 Wochen der Studie 190-201 beendet hatten. Somit wurden alle 23 Patienten der ITT-Population der Studie 190-201 in die Extensionsstudie aufgenommen, auch wenn a priori festgelegt wurde, dass diejenigen Patienten, die einen Abfall des ML-Scores um ≥ 3 Punkte während der Studie 190-201 (von Studienbeginn bis Studienende) aufwiesen, ausgeschlossen werden sollten. Die Rationale für einen Cut-off-Wert von ≥ 3 Punkte ist von Seiten des pU nicht begründet worden, traf jedoch auch auf keinen Patienten zu.

Das mediane Alter der insgesamt 24 Patienten bei Studienschluss der Studie 190-201 lag bei 4 Jahren (Spanne: 3–8 Jahre). Die Diagnosestellung lag im Median bei 3 Jahren (Spanne: 2,5–6,3 Jahre). Die Mehrheit der Studienteilnehmer war weiblich (63 %) und Kaukasier (96 %). Insgesamt 20 der 24 eingeschlossenen Patienten hatten zur Screening-Visite der Studie 190-201/190-202 einen ML-Score von 3–4 Punkten und entsprechend einen ML-Score im unteren protokollgemäßen Bereich (Einschlusskriterium: ML-Score von 3–6 Punkten). Zum Baseline-Zeitpunkt der Wirksamkeitsevaluation (Beginn der 300-mg-Dosisgabe) hatten drei Patienten einen ML-Sore < 3 Punkte (n=1 ML-Score von 1; n=2 ML-Score von 2).

Insgesamt wurden fünf Änderungen am Studienprotokoll (erstes Protokoll vom 07. März 2013) der Studie 190-201 vorgenommen. Das finale Protokoll (Version 5) wurde am 24. September 2014 angefertigt, während der erste Patient bereits am 13. September 2013 in die Studie aufgenommen wurde. Wichtige Änderungen umfassten u. a. die Möglichkeit der Gabe von Antihistaminen und die Möglichkeit der Gabe zusätzlicher Antipyretika (ca. 30 min vor der Infusion von Cerliponase alfa) aufgrund von möglichen Überempfindlichkeitsreaktionen durch die ERT-Gabe (Amendement 1 vom 10. Mai 2013) und die Festlegung, dass der SAP in seiner aktuellen Version 3 (Stand: 26. März 2016) gilt, sollte es Diskrepanzen zwischen der Beschreibung der statistischen Analyse im Protokoll und dem finalen SAP geben (Amendement 2 vom 26. Juli 2013). Für die Studie 190-202 gab es insgesamt fünf Protokolländerungen, wovon die ersten drei den vorliegenden aktuellen Datenschnitt vom 03. Juni 2016 betreffen (Amendement 3 vom 26. Februar 2016).

Die Fallzahlplanung und statistischen Auswertungen der Studie 190-201/190-202 sind im SAP beschrieben, der für beide Studien 190-201 und 190-202 gleichermaßen gilt (siehe Kapitel 2.5.4.1). Gemäß Studienbericht 190-202 kam es zur Änderung der Zielsetzung zwischen dem Protokoll 190-202 und SAP 190-201/202, u. a. wurde der Endpunkt Lebensqualität, zuvor im Protokoll als sekundärer Endpunkt spezifiziert, im SAP als explorativer Endpunkt festgelegt (bei Diskrepanzen galten die Angaben im SAP). Die erste Version des SAP ist auf den 03. Mai 2015 datiert und wurde somit post-hoc nach Einschluss des ersten Patienten in die Studie 190-201 (13. September 2013) verfasst. Der finale SAP (Version 3) ist auf den 25. März 2016 datiert und liegt somit 4 Monate nach der letzten Visite der Studie 190-201 (30. November 2015). Inwieweit möglicherweise Änderungen zur Datenauswertung in Kenntnis der Daten festgelegt wurden, ist nicht nachzuvollziehen. So ist in der letzten Version des Studienprotokolls (Amendement 5, vom 24. September 2014) ausgeführt, dass im SAP festgelegt sei, dass sowohl die beiden Subskalen motorische Fähigkeiten und Sprachfähigkeiten als auch die Gesamtskalen des Hamburg- und WCMC-Instruments mit den historischen Kontrolldaten verglichen würden. Dies ist im finalen SAP jedoch nicht enthalten.

Die dargestellten und ausgewerteten Daten der Studie 190-201 beziehen sich auf den Studienbericht vom 27. April 2016 (Studienende: 30. November 2015). Für die Studie 190-202

liegen zwei nicht a priori definierte Analysen mit Datenschnitten vom 15. Oktober 2015 (keine Angabe zur Veranlassung der Zwischenanalyse) und vom 03. Juni 2016 vor, wobei die Analyse mit Datenschnitt vom 03. Juni 2016 auf Ersuchen der FDA durchgeführt wurde. Zum Zeitpunkt des Datenschnitts vom 03. Juni 2016, der für die vorliegende Nutzenbewertung als bewertungsrelevant herangezogen wurde, betrug die Exposition mit einer Dosis von 300 mg bei den N=23 Patienten der ITT-Population mindestens 72 Wochen.

Das unkontrollierte Design beider Studien erlaubt keine Berechnung vergleichender Therapieeffekte, die eine valide und unverzerrte Einschätzung der Behandlung mit Cerliponase ermöglichen würden. Eine abschließende Aussage über Wirksamkeit und Sicherheit der Intervention kann allein anhand der Studien 190-201/202 nicht getroffen werden.

3.3 Design und Methodik der Studie 190-901 und des historischen Vergleichs

Gemäß dem finalen SAP (Version 3, 25. März 2016) der Studie 190-201/190-202 war für die primäre Auswertung ein Vergleich mit einer historischen Kontrollgruppe vorgesehen. Dieser wurde auf Empfehlung der Zulassungsbehörden im Statistischen Analyseplan der „Integrated Summary of Efficacy“ (SAP ISE) spezifiziert: „Auf Empfehlung der Zulassungsbehörden wurde im ISE SAP eine Responder-Analyse nach 1:1-Paarbildung (Matching) der Teilnehmer der Studie 190-201 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) als primärer Endpunkt festgelegt“ [2]. Die aktualisierte Analyse erfolgte auf Basis des DEM-CHILD-Datensatzes (Supplementary Report der Studie 190-901).

Das DEM-CHILD-Register stellt eine Erweiterung des ursprünglichen Längsschnittdatensatzes (n=30) aus Hamburg dar. Bei diesem Register handelt es sich um eine fortlaufende, multizentrische und multinationale, klinische, europaweite Datenbank mit Sitz in Hamburg. Sie enthält, verglichen mit der Original-Längsschnittdatenbank, erweiterte und aktualisierte klinische Daten. Zum Zeitpunkt der ersten Analyse des historischen Vergleichs (30. November 2015) lagen verfügbare Daten von N=74 CLN2-Patienten aus zwei Zentren des DEM-CHILD-Registers (Hamburg, n=63, und Verona, n=11) mit Datenschnitt vom Februar 2015 vor. Von den 74 CLN2-Patienten verblieben nach Einsatz der HML-Filter und Ausschluss eines Zwillings 41 Patienten aus dem Register (55 %; auswertbare Patienten) für eine weitergehende Analyse. Diese Patienten erfüllten die Eignungskriterien für einen historischen Vergleich mit den Patienten der Studie 190-201/202 (nähere Informationen zum Selektionsprozess siehe 2.4.2). Von den 41 Patienten, die die Eignungskriterien für den historischen Vergleich erfüllten, handelte es sich um 26 Patienten aus dem ursprünglichen Längsschnittdatensatz aus Hamburg (n=30). Allerdings wurden von den ursprünglichen 30 Patienten für die Analyse zur Eignung als Kontrollen nur 29 der 30 Patienten ausgewertet. Es bleibt unklar, ob die Teilpopulation von 26 Patienten der insgesamt 41 eingeschlossenen Patienten des DEM-CHILD-Registers sich auf die 30 Patienten oder die 29 ausgewerteten Patienten des ursprünglichen Längsschnittdatensatzes bezieht bzw. warum ein Patient nicht herangezogen wurde. Weitere wichtige Aspekte und Limitationen der vorhandenen Angaben betreffen die Vollständigkeit, Repräsentativität und Selektion der möglichen Kontrollen für den historischen Vergleich. DEM-CHILD ist ein internationales Register, trotzdem wurden nur Patienten aus den zwei Zentren Hamburg und Verona ausgewählt. Es ist unklar, ob die Grundgesamtheit des gesamten DEM-CHILD-Registers repräsentativ für die CLN2 ist oder ob eher schwerere CLN2-Verläufe eingeschlossen werden und somit die historischen Kontrollen stärker betroffen sind. Für den historischen Vergleich wird zwar hinsichtlich des HML-Scores gematcht,

jedoch bildet diese motorische und sprachliche Subskala nur bestimmte Krankheitsaspekte ab. Desweiteren kann mangels Angaben nicht beurteilt werden, inwieweit auch die Teilgruppe aus den lediglich zwei Zentren (Hamburg und Verona) repräsentativ für das Krankheitsbild ist und inwieweit diese Auswahl zu Selektionseffekten, gerade auch hinsichtlich krankheitsrelevanter prognostischer Faktoren führt. Das gleiche gilt für die Auswahl der für das eigentliche Matching zur Verfügung stehenden 41 bzw. 42 möglichen Kontrollen aus den ursprünglich verfügbaren N=74 Patienten aus dem Register. Allein 16 Patienten mussten wegen nicht vorhandener HML-Werte ausgeschlossen werden (siehe Tabelle 3).

Im ISE SAP wurde das Analyseverfahren nach 1:1-Paarbildung festgelegt (1:1-Matching der N=23 ITT-Patienten aus Studie 190-201/202 mit historischen Kontrollen aus dem Pool der N=41 bzw. N=42 passenden Kontrollen aus dem DEM-CHILD-Register). Der pU legt im Modul 4 die erste Analyse vom 30. November 2015 vor. Die zugrunde liegenden Daten umfassen den Datenschnitt vom 15. Oktober 2015 der Studie 190-201/190-202 [N=23] und den Datenschnitt vom Februar 2015 der Studie 190-901 [N=41, siehe oben]. Als Begründung verweist der pU im Modul 4 darauf, dass nur diese eine Vergleichbarkeit der Gruppen hinsichtlich prognostisch relevanter Faktoren gewährleisten [2]. Zusätzlich liegen aktuellere, nicht a priori festgelegte Analysen vom 15. Juni 2016 und vom 01. November 2016 vor, wobei die Analyse vom 01. November auf Ersuchen der FDA durchgeführt wurde. In diese Analysen werden für die Studie 190-201/190-202 der aktuellste Datenschnitt vom 03. Juni 2016 [N=23] und für die Studie 190-901 [N=42] vom August 2016 herangezogen. Somit wurde in den Pool der Patienten für die historische Kontrollgruppe ein weiterer Patient aufgenommen (auswertbare Population der Studie 190-901 [N=41, Datenschnitt Februar 2015] vs. [N=42, Datenschnitt August 2016]). Es stellt sich die Frage, warum sich die Analyse der Studie vom 15. Juni 2016 auf zeitlich spätere Daten der historischen Kontrollgruppe bezieht, da diese erst im August 2016 transferiert bzw. ausgewertet wurden. Darüber hinaus wird im Studienbericht vom 15. Juni 2016 erstmalig berichtet, dass Patienten aus dem DEM-CHILD-Register in eine Cerliponase-alfa-Studie eingeschlossen wurden (N=7). Um welche Studie es sich konkret handelt, wird nicht beschrieben. Somit wurden zum Datenschnitt der historischen Kontrollgruppe im August 2016 insgesamt 49 Patienten als auswertbare Population identifiziert, von denen allerdings 7 Patienten ausgeschlossen wurden, da sie an einer Cerliponase-alfa-Studie teilnahmen, so dass der Pool von N=42 Patienten für das Matching resultierte. Es ist unklar, inwieweit der zusätzliche Patient (N=42) das 1:1-Matching beeinflusst haben könnte, da hierzu im Studienbericht keine Angaben gemacht wurden. Es ist nicht klar, ob dieser eine zusätzliche Patient für das 1:1-Matching der aktuelleren Datenanalysen ausgewählt wurde und ggf. einen anderen Patienten der ersten Analyse des historischen Vergleichs ersetzte. Daher bleibt unklar, inwieweit diese Änderungen die Ergebnisse beeinflusst haben könnten, zumal der pU im Modul 4 angibt, dass nur der erste Datenschnitt eine Vergleichbarkeit der Gruppen hinsichtlich prognostisch relevanter Faktoren gewährleistet. Im Dossier sind keine aktualisierten Daten hinsichtlich soziodemographischer und krankheitsbezogener Charakteristika angegeben (siehe Tabelle 13). Ausgehend von der Annahme eines identischen Vorgehens beim Matching wurden trotz dieser Limitationen für die Nutzenbewertung die Ergebnisse der aktuelleren Analyse vom 15. Juni 2016 extrahiert (190-901: Population möglicher historischer Kontrollen N=42, Datenschnitt August 2016 und 190-201/190-202: ITT N=23, Datenschnitt 03. Juni 2016). Für die Ereigniszeitanalysen des Endpunkts „stabiler Abfall um ≥ 2 Punkte“ der ML-/HML-Skala wurde die spätere Analyse vom 1. November 2016 mitberücksichtigt. Der Vergleich der soziodemographischen und krankheitsbezogenen Charakteristika muss sich allerdings auf die vom pU im Modul 4 dargestellten Angaben der Datenanalyse vom 30. November 2015 (190-901: Population möglicher historischer

Kontrollen N=41, Datenschnitt Februar 2015 und 190-201/190-202: ITT N=23, Datenschnitt 15. Oktober 2015) beziehen.

Unterschiede zwischen den 41 auswertbaren Registerpatienten und den 23 Patienten der Studien 190-201/202 bestehen – soweit Angaben vorhanden – hauptsächlich hinsichtlich des mittleren Alters bei Diagnose (41 Monate in Studie 190-201 und 60 Monate im DEM-CHILD-Register) sowie hinsichtlich der historisch-zeitlichen Erhebungszeiträume und der Ausgangswerte der Morbiditätsdaten zu Motorik und Sprache [(H)ML-Score]. Das früheste Geburtsjahr in Studie 190-201 ist 2005 (Einschlussdatum erster Patient minus Maximalalter bei Einschluss), im Register liegt es in der Spanne von 1960–1969, mit mehrheitlich zwischen 1990–1999 (46 %) und 2000–2009 (39 %) geborenen Kindern. Der mittlere (H)ML-Score ist nicht direkt vergleichbar, da zeitlich unterschiedlich erhoben: In Studie 190-201 lag der mittlere ML-Score bei Studieneinschluss bei 3,7 Punkten (bei einem mittleren Alter von 4,3 Jahren). Im DEM-CHILD-Register lag der mittlere HML-Score bei 3,0 Punkten (siehe Tabelle 13; eigene Berechnungen zu 29 Kindern mit vorhandenen Angaben), allerdings vor oder zum Zeitpunkt der Diagnosestellung [mittleres Alter bei Diagnose 5 Jahre]).

Analysiert wurden die Patienten der ITT-Population der Studie 190-201/202, denen Patienten der historischen Kontrolle im Verhältnis 1:1 zugeteilt wurden (Methode #1), vorausgesetzt, der ML-/HML-Wert zu Baseline stimmte exakt überein und der Altersunterschied war so gering wie möglich (allerdings nicht größer als 12 Monate). 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. Bei einem der 23 Patienten der Studie 190-201 war eine Paarbildung nicht möglich, da der Altersunterschied zu groß war (21 Monate statt maximal 12 Monate).

In der gematchten Auswertungspopulation des historischen Vergleichs betrug der mittlere (SD) Altersunterschied zwischen der Cerliponase-alfa-Gruppe und der historischen Kontrollgruppe 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. Die Geschlechterverteilung zwischen den Gruppen war jedoch nicht ausgeglichen. Insgesamt 16 der 22 Patienten (68 %) der historischen Kontrollgruppe (190-901) und 9 der 22 Patienten (41 %) der Studie 190-201/190-202 waren männlich. Hinsichtlich des Genotyps wurden bei insgesamt 14 der 22 Patienten (64 %) der Studie 190-901 verglichen mit 9 von 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 190-201/190-202 als der Studie 190-901 („matched“) zwei der eher selten auftretenden Allele auf (6 Patienten [28 %] 190-201/202, 4 Patienten [18 %] 190-901). Für die gematchte Analysepopulation kann hinsichtlich der oben beschriebenen Unterschiede der historisch-zeitlichen Erhebungszeiträume (unterschiedliche Geburtsjahrgänge) keine Aussage gemacht werden, da dazu keine Daten vorliegen. Inwieweit sich diese zeitlichen Unterschiede durch Faktoren wie z. B. unterschiedliche Aufmerksamkeit oder diagnostische Kriterien für die Erkrankung oder unterschiedliche Erfassungsmodalitäten der CLN2-Morbiditätsdaten verzerrend auf die Ergebnisse des historischen Vergleichs auswirken, kann nicht abschließend eingeschätzt werden.

Es ist unklar, warum in der historischen Kontrollgruppe (siehe Tabelle 13) keine Patienten einen HML-Score von 6 hatten, im Rahmen des Matchings (190-201/190-202 vs. 190-901) jedoch zwei Personen einen Score von 6 aufwiesen (siehe Tabelle 14). Ob dies eine mögliche Selektion hin zu schwereren Krankheitsstadien bedeutet, kann aufgrund dieser Unklarheit nicht beurteilt werden.

Die Geschlechterverteilung zwischen den Gruppen war nicht ausgeglichen: Insgesamt 32 % der historischen Kontrollgruppe (190-901) gegenüber 59 % der Studie 190-201/190-202 waren weiblich.

Die nachfolgend berichteten Ergebnisse beziehen sich auf den Datenschnitt vom 15. Juni 2016. Dabei beziehen sich die Daten der Studie 190-201/190-202 auf den Datenschnitt vom 03. Juni 2016 (die 23 Patienten der ITT-Population waren bis zu diesem Zeitpunkt über mehr als 72 Wochen behandelt worden). Zu Woche 89 liegen für diesen Datenschnitt nur Daten von 11 Patienten der historischen Kontrollgruppe und von 13 Patienten in der Cerliponase-alfa-Gruppe vor (siehe Tabelle 24). Nach Verfügbarkeit werden die Daten aus dem aktuellsten Datenschnitt vom 01. November 2016 mit 21 gematchten Paaren ergänzt dargestellt (nur für primäre Responderanalyse und Ereigniszeitanalyse). Bis zu diesem Zeitpunkt waren alle Patienten über mindestens 96 Wochen mit Cerliponase alfa in 300-mg-Dosierung behandelt worden. Für diesen aktuellsten Datenschnitt liegt ein Matching-Paar weniger vor. Da weder Angaben zur Auswahlprozedur bzw. den Gründen für die geringere Anzahl an gebildeten Paaren noch die Patientencharakteristika bezüglich der soziodemographischen, prognostischen und weiteren krankheitsrelevanten Faktoren vorliegen, kann die Güte des historischen Vergleichs nicht ausreichend eingeschätzt werden. Zudem sind keine Angaben ersichtlich zur Vollständigkeit bzw. zu fehlenden Messwerten der Skalen zu den einzelnen Messzeitpunkten. Wie oben beschrieben lagen für den vorigen Datenschnitt zum 15. Juni 2016 für den Beobachtungszeitraum bis Woche 89 nur für rund 50 % der Patienten aus beiden Vergleichsgruppen Skalenwerte vor. Wie sich dies für den Datenschnitt vom 1. November 2016 bzw. für den Beobachtungszeitraum 96 Wochen darstellt, ist nicht ersichtlich. Insgesamt muss als weitere Limitation hinsichtlich der Ergebnissicherheit des historischen Vergleichs angemerkt werden, dass nicht ersichtlich wurde, anhand welcher „Endwerte“ die Linie des Abfalls des ML-HML-Scores geschätzt wurde. Da das Ausmaß des Abfalls (Slope) lediglich anhand eines Anfangswerts (zu Baseline) und des jeweils vorhandenen „Endwerts“ bestimmt wurde, wäre zur Beurteilung dieses Verfahrens eine zeitliche Übersicht mit Messwerten wichtig: d. h. die Kenntnis der jeweiligen Messzeitpunkte dieser „Endwerte“ bzw. die resultierende jeweilige Beobachtungszeit bis zu diesem „Endmesswert“ des ML-HML-Scores.

Zur Erhebung der Morbidität aufgrund der CLN2-Erkrankung wurde in der historischen Kontrollgruppe die ursprünglich entwickelte ‚HML-Skala‘ (kombinierte Skala aus motorischer und sprachlicher Domäne) verwendet. Diese HML-Skala wurde, wie in Abschnitt 3.2 und Tabelle 6 beschrieben, für die einarmigen Studien 190-201/190-202 mit den Entwicklern der HML-Skala marginal angepasst (‚ML-Skala‘). Diese Anpassungen erscheinen für eine Verwendung der Skala – gerade auch in einer Studie – nachvollziehbar und sinnvoll. Es kann angenommen werden, dass die wahrscheinlich etwas niedrigere Genauigkeit und Diskriminationsfähigkeit der ursprünglichen, auch im Register verwendeten HML-Skala sich nicht maßgeblich verzerrend auf die Ergebnisse auswirkt. Es ist jedoch schwierig, dies abschließend zu beurteilen. Bedeutender für eine mögliche Ergebnisunsicherheit ist jedoch, dass im Register für die historischen Kontrollen diese HML-Morbiditätsdaten teilweise retrospektiv erhoben wurden. Da hierzu keine weiteren Erläuterungen und Diskussionen seitens des pU vorliegen, kann das Ausmaß der möglichen Verzerrungen diesbezüglich nicht beurteilt werden.

Zusammenfassende Beurteilung des historischen Vergleichs

Aufgrund des Fehlens einer Vergleichstherapie führt der pU einen historischen Vergleich zur Ableitung des Zusatznutzens von Cerliponase alfa durch, allerdings nur für die Responder-, Slope-

und Ereigniszeitanalyse der ML- bzw. HML-Skala. Für die Endpunktkategorien Mortalität, Lebensqualität und Nebenwirkungen ist aufgrund fehlender Daten in der historischen Kontrollgruppe kein Vergleich möglich.

Zentrale Voraussetzungen für einen historischen Vergleich sind sowohl Vollständigkeit als auch mögliche gleiche Erfassungsmethode der zugrundeliegenden Daten, vor allem hinsichtlich prognostischer Faktoren, Repräsentativität und Selektion der möglichen Kontrollen für den historischen Vergleich. Es bleibt unklar, ob die Grundgesamtheit des gesamten DEM-CHILD-Registers und auch die Auswahl der möglichen Kontrollen repräsentativ für die CLN2-Erkrankung ist oder welche Verzerrungen aufgrund von Selektionseffekten aufgetreten sein können und ob insbesondere eher schwerere CLN2-Verläufe eingeschlossen wurden und somit die historischen Kontrollen stärker betroffen sind. Für die Ergebnissicherheit eines historischen Vergleichs ist die Vollständigkeit der Informationen und die hinreichende Übereinstimmung der Charakteristika der betrachteten Studienpopulationen mit der Zielpopulation, vor allem hinsichtlich Risiko- bzw. prognostischen Faktoren, Vortherapie(n) und Therapiedauer, ganz wesentlich. Da für die historische Kontrollgruppe keine Informationen zu früheren Therapien, Begleitmedikationen oder Vorerkrankungen vorliegen, ist eine Beurteilung der Vergleichbarkeit der beiden Gruppen hierzu nicht möglich. Anhand der limitierten vorliegenden Patientencharakteristika lassen sich nur begrenzt Rückschlüsse auf die Vergleichbarkeit ziehen. Beispielsweise war die Geschlechterverteilung zwischen den gematchten Gruppen nicht ausgeglichen und es wiesen mehr Patienten der Studie 190-201/190-202 zwei der eher selten auftretenden Allele auf (6 Patienten [28 %] 190-201/202, 4 Patienten [18 %] 190-901). Darüber hinaus wurde der HML-Score im Register teilweise retrospektiv erhoben und die (H)ML-Scores aus beiden Studien historisch-zeitlich unterschiedlich erhoben. Inwieweit sich diese teilweise retrospektive Erhebung und diese zeitlichen Unterschiede durch Faktoren wie z. B. unterschiedliche Aufmerksamkeit oder diagnostische Kriterien für die Erkrankung oder unterschiedliche Erfassungsmodalitäten der CLN2-Morbiditätsdaten verzerrend auf die Ergebnisse des historischen Vergleichs auswirken, kann nicht abschließend eingeschätzt werden.

Wie oben bereits adressiert, wurden in der aktuellen Analyse vom 15. Juni 2016 insgesamt 42 auswertbare Patienten anstelle der zuvor 41 Patienten herangezogen. Es bleibt unklar, inwieweit der zusätzliche Patient das 1:1-Matching beeinflusst haben könnte, da hierzu im Studienbericht keine Angaben gemacht wurden. Zudem ist fraglich, ob sich hinsichtlich soziodemographischer und krankheitsbezogener Charakteristika Änderungen hätten ergeben können, die das 1:1-Matching beeinflusst hätten. Ebenso bleibt unklar, wie es für den aktuellsten Datenschnitt vom 1. November 2016 zu den 21 Matching-Paaren kam und wie deren Verteilung hinsichtlich der prognostischen Faktoren ist.

3.4 Wirksamkeit

Zur Wirksamkeit werden für die einarmige Studie 190-201/190-202 die Responder-, Slope- und Ereigniszeitanalyse der ML-Skala sowie die Veränderung der Punktzahl auf der ML-Skala bzw. MLV(S)-Skalen im Vergleich zu Baseline dargestellt.

Zur Wirksamkeit für den 1:1 gematchten historischen Vergleich (Studie 190-201/190-202 [N=22] vs. Studie 190-901 [N=22] liegen Daten der Responder-, Slope- und Ereigniszeitanalysen für die ML-/HML-Skala vor.

Soweit nicht anders angegeben, beziehen sich die Ergebnisse auf Baseline zur 300-mg-Dosis, d. h. die letzte Beobachtung vor ersten der 300-mg-Infusion.

Ergebnisse der einarmigen Studien 190-201/190-202 (03. Juni 2016)

Todesfälle

Das Gesamtüberleben wurde als UE für die Sicherheitspopulation erfasst: Bis zum aktuellen Datenschnitt (03. Juni 2016) traten keine Todesfälle auf.

Responderrate der ML-Skala (primärer Endpunkt, Responderanalyse)

Zu Woche 48 (Studientag 340 bezogen auf die erste Infusion von 300 mg in der Studie 190-201) und zu Woche 80 (Studientag 540 bezogen auf die erste Infusion von 300 mg in der Studie 190-201) blieb die Rate der Responder unverändert (87 %).

Abfall der ML-Skala skaliert auf 48 Wochen (Slopeanalyse, ursprünglicher primärer Endpunkt)

Der mittlere Abfall der ML-Skala betrug 0,4 Punkte [95%-KI 0,05; 0,75] in der Studie 190-201 und 0,32 [95%-KI 0,13;0,52] in der Studie 190-201/190-202.

Veränderung der ML- bzw. MLV(S)-Skalen im Vergleich zu Baseline

Bei der ML-Skala und MLV-Skala verschlechterten sich die Werte bis Woche 73, im Durchschnitt um 0,4 Punkte und um 0,6 Punkte, während sie bei der MLVS-Skala unverändert blieben.

Ergebnisse für den historischen Vergleich (190-201/190-202 vs. 190-901; Analyse vom 15. Juni 2016)

Todesfälle

Da in der historischen Kohortenstudie 190-901 (DEM-CHILD) das Gesamtüberleben nicht erfasst wurde, ist ein Vergleich mit der Interventionsstudie nicht möglich.

Responderanalyse der ML-/HML-Skala (primärer Endpunkt)

Die Definition des Endpunkts Response der ISE differiert von der Responderdefinition in der einarmigen Studie 190-201/202: Für den historischen Vergleich der ISE ist er definiert anhand der Größe des linearen Abfalls der motorischen und sprachlichen Fähigkeiten, gemessen mittels der (H)ML-Werte. Responder waren somit definiert als Patienten, bei denen der Abfall (Slope der Linie), skaliert auf 48 Wochen, kleiner als zwei (< 2) war. Die Festlegung dieses Schwellenwerts beruhte auf Längsschnittdaten zum natürlichen Krankheitsverlauf bei 29 unbehandelten Patienten aus der Hamburger Registerkohorte. Bei dieser Kohorte fiel der ML-/HML-Score nach Auftreten erster Symptome innerhalb von 18 Monaten im Mittel um zwei Punkte (SD von 1,8) ab. Dieses Kriterium wurde im Rahmen der Primäranalyse der Studie 190-901 auf einen Zeitraum von 12 Monaten (48 Wochen) festgelegt. Es ist unklar, weshalb für die Responderanalyse anhand dieser Verlaufsdaten der Zeitraum des Abfalls um 2 Punkte auf 48 Wochen verkürzt wurde und ob die Schwellenwerte von 2 im Sinne eines minimalen klinisch relevanten Unterschieds zu interpretieren sind. Die patientenindividuellen Daten der Hamburger Kohorte zeigen einen starken Abfall auf der kombinierten Motor- und Sprach-HML-Skala. Der mittels linearer Regression geschätzte Abfall der aggregierten Verläufe ergab eine Steigung (Slope) bzw. Abfall der motorischen und sprachlichen Fähigkeiten um 2,18 Punkte pro Jahr (siehe Abbildung 7 und Abbildung 8 im Anhang). Aufgrund dieser kleinen Hamburger Stichprobe erscheint der Schwellenwert für die Responderanalyse zumindest aufgrund dieser Schätzung anhand der klinischen HML-Werte nachvollziehbar.

Inwieweit diese Kohorte repräsentativ für den natürlichen Verlauf der Erkrankung ist, kann anhand der Daten jedoch nicht beurteilt werden.

Das Ansprechen (Größe des Abfalls des ML-/HML-Scores um < 2 Punkte bezogen auf 48 Wochen) betrug 100 % (n=22/22) bei den Patienten der Studie 190-201/190-202 verglichen mit 45 % (n=10/22) in der historischen Kontrollgruppe (p=<0,0001). Dieses Ergebnis bestätigt sich in der Analyse vom 01. November 2016 (wonach die Patienten mindestens 96 Wochen unter Behandlung waren) und durch die supportiven Populationen (Methode #1 [Simulation] und Methode #2).

Abfall der ML-Skala skaliert auf 48 Wochen (Slopeanalyse)

Der ML-/HML-Score skaliert auf 48 Wochen fiel bei den Patienten der Studie 190-901 im Mittel um 2,0 Punkte ab, verglichen mit 0,34 Punkten bei den Patienten der Studie 190-201/202, was einem Unterschied von 1,66 Punkten (p<0,0001) entspricht. Dieses Ergebnis wird ebenfalls durch die supportive Population (Methode #2) bestätigt.

Die Ergebnisse der Sensitivitätsanalysen zur Response und zum Slope zeigen ähnliche und gleichgerichtete Effekte zu denen der Primäranalyse.

Ereigniszeitanalysen

Die Ergebnisse zur a priori definierten Kaplan-Meier-Analyse (Datenschnitt: 01. November 2016, adjustiert nach ML-/HML-Score und Alter jeweils zu 300-mg-Baseline), ausgedrückt als Zeit, bis der ML-/HML-Score nachhaltig um ≥ 2 Punkte oder auf den Wert von Null abfiel, zeigen ein Hazard Ratio (HR) von 0,10 ([95%-KI 0,03; 0,38], p=0,0005). Die mediane Zeit bis zur Krankheitsprogression betrug bei den Patienten der historischen Kontrolle 285 Tage (95%-KI: 210; 420), während der Median bei den behandelten Patienten noch nicht erreicht ist. Der Unterschied der Ereignisraten war zwischen den Gruppen in Woche 48 signifikant und stieg bis Woche 72 und 96 weiter an. Der Gruppenunterschied war in Woche 96 signifikant: HR=0,05 ([95%-KI 0,01; 0,17], p<0,0001).

Somit ergibt sich eine relative Risikoreduktion einer Krankheitsprogression, gemessen anhand des linear geschätzten Verlaufs der ML-Skala, von 90 % bei Patienten unter Cerliponase alfa im Vergleich zu Patienten der historischen Kontrolle. Allerdings wurde nicht diskutiert, ob die Voraussetzungen für die Anwendung eines Cox-Modells, z. B. proportional hazards, Gründe und Häufigkeiten von Zensierungen, gegeben sind.

Veränderung auf den ML- bzw. MLV(S)-Skalen im Vergleich zu Baseline

Der ML-/HML-Score nahm bei den unbehandelten Patienten der historischen Kontrolle im Mittel zu Woche 73 um -2,7 Punkte (N=22) ab, verglichen mit -0,5 Punkte (N=22) bei den behandelten Patienten. Zu Woche 89 liegen nur Daten von 11 Patienten der historischen Kontrollgruppe und von 13 Patienten in der Cerliponase-alfa-Gruppe vor. Zur MLV- und MLVS-Skala liegen nur Ergebnisse einer nicht-gematchten Analyse vor, so dass diese nicht berücksichtigt wurden.

Zusammenfassende Einschätzung zur Wirksamkeit

Das unkontrollierte Design beider Studien 190-201/190-202 erlaubt keine Berechnung vergleichender Therapieeffekte, die eine valide und unverzerrte Einschätzung der Behandlung mit Cerliponase ermöglichen würden. Eine abschließende Aussage über die Wirksamkeit der Intervention kann deshalb allein anhand der Daten aus den Studien 190-201/190-202 nicht getroffen werden.

Der pU legt zusätzlich einen Vergleich mit historischen Kontrollen aus einem Register vor. Die Datenbasis und die Auswahl der historischen Kontrollen für das 1:1-Matching weisen vielfältige Limitationen auf (siehe „Zusammenfassende Beurteilung des historischen Vergleichs“ in Abschnitt 3.3), die die Validität der Ergebnisse einschränken.

Die Ergebnisse des historischen Vergleichs (190-201/190-202 vs. 190-901; Analyse vom 15. Juni 2016) zeigen für die Responder-, Slope- und Ereigniszeitanalysen statistisch signifikante Effekte zugunsten von Cerliponase alfa. Die Sensitivitätsanalysen zur Response und zum Slope zeigen ähnliche und gleichgerichtete Effekte wie die Ergebnisse der primären Analyse. Die Ergebnisse des historischen Vergleichs liegen jedoch bis auf die Ereigniszeitanalysen nicht in einer Größenordnung, mit der ausgeschlossen werden könnte, dass die beobachteten Unterschiede allein auf systematischer Verzerrung beruhen. Angesichts der zahlreichen im Abschnitt 3.3 beschriebenen methodischen Limitationen des historischen Vergleichs und der viel geringeren Effektstärke als bei der primären Responderanalyse bleibt auch das Ergebnis der Ereigniszeitanalyse mit höherer Ergebnisunsicherheit verbunden.

3.5 Lebensqualität

Die Lebensqualität wurde in der Studie 190-201/190-202 mit dem generischen PedsQL™-Fragebogen, dem ‚Parent Report for Toddlers‘ erfasst. Die Ergebnisse, Veränderung von Studien-Baseline bis Studienende (Woche 49), werden für die einarmige Studie 190-201/190-202 deskriptiv berichtet, da keine Vergleichsdaten aus der historischen Kohorte 190-901 zur Verfügung stehen. Für die Studie 190-201 zeigte sich in der ITT-Population (N=23) eine durchschnittliche Verbesserung im Gesamtscore um 2,6 Punkte. Gleichermaßen zeigte sich bis auf die Dimension ‚Körperliche Kompetenzen‘ (Verschlechterung im Mittel von 6,1 Punkten) in allen weiteren Dimensionen ebenso wie im psychosozialen Gesamtscore (+7,9 Punkte) eine Verbesserung der Lebensqualität.

Zum Datenschnitt 03. Juni 2016 waren von 12 Patienten (52 %) der Studie 190-201/190-202 Angaben bis Woche 98 vorhanden. In allen Dimensionen zeigte sich im Vergleich zu Baseline eine Verschlechterung. So reduzierte sich der Gesamtscore im Mittel um 14,1 Punkte, die stärkste Verschlechterung zeigte sich in den Körperlichen Kompetenzen (Abnahme im Mittel um 27,9 Punkte).

Zusammenfassende Einschätzung zur Lebensqualität

Die Veränderung der Lebensqualität wurde in der Studie anhand des generischen, validierten PedsQL™-Moduls ‚Parent Report for Toddlers‘ erfasst. Hierbei handelt es sich um ein Fremdbeurteilungsinstrument zur Erfassung der Lebensqualität bei pädiatrischen Populationen im Alter von 2 bis 4 Jahren mit chronischen Erkrankungen mit ausreichender Testgüte. Das mediane Alter bei Einschluss in die Studie 190-201 lag bei 4 Jahren (Alterspanne 3–8 Jahre). Es bleibt unklar, weshalb der pU für Kinder höherer Altersklassen in der Studie nicht andere bestehende Versionen, wie bspw. Versionen für Vorschulkinder (5 bis 7 Jahre) und Schulkinder (8 bis 12 Jahre), eingesetzt hatte. Der eingesetzte PedsQL 4.0 Generic Core Scales ist jedoch kein krankheitsspezifisches Instrument und kann keine spezifischen Aspekte adressieren, die die Lebensqualität bei pädiatrischen Patienten mit neurodegenerativen Erkrankungen beeinflussen. Da keine vergleichenden Daten und keine Informationen zur klinischen Relevanz einer Veränderung vorliegen, können für die Bewertung des Zusatznutzens keine validen und

abschließenden Aussagen getroffen werden. Der PedsQL™ scheint möglicherweise andere Aspekte der Erkrankung bei den Kindern bzw. Beurteilern/Angehörigen abzubilden als die ML-Skala. Die Kinder scheinen sich im PedsQL™ zur Analyse am 3. Juni 2016 deutlicher zu verschlechtern als im ML-Score. Die stärkste Verschlechterung zeigte sich im PedsQL™ in den Körperlichen Kompetenzen; diese nahmen im Mittel um 27,9 Punkte ab.

3.6 Sicherheit

Zur Beurteilung von sicherheitsrelevanten Endpunkten diente die Sicherheitspopulation, die alle Patienten umfasste, denen ein ICV-Port implantiert worden war (N=24). Die hier berichteten Ergebnisse bilden den Datenschnitt vom 3. Juni 2016 der Folgestudie 190-202 ab (gesamtes Dosisintervall einschließlich der Studie 190-201). Bei allen in der Studie 190-201/190-202 mit einer Dosis von 300 mg behandelten Patienten betrug die Exposition mindestens 72 Wochen. Während der Studie 190-201 betrug die mittlere Dauer der Exposition mit Cerliponase alfa 47,1 Wochen in der 300-mg-Dosierungsperiode und für die Studie 190-201/190-202 im Median 95 Wochen zum Datenschnitt 03. Juni 2016.

Unerwünschte Ereignisse

Insgesamt traten bei allen Studienteilnehmern unerwünschte Ereignisse auf. Kein UE führte zu einem Abbruch der Studienmedikation oder zum Tod. Die am häufigsten berichteten UEs waren (Studie 190-201/Studie 190-201/202): Fieber (54 %/71 %), Krampfanfall (beide 58 %), Erbrechen (46 %/58 %), Infektion der oberen Atemwege (42 %/50 %), generalisierte tonisch-klonische Anfälle (8 %/46 %), Epilepsie (beide 46 %) und Rhinitis (21 %/42 %).

Unerwünschte Ereignisse des NCI-CTCAE-Grades 3 und höher

Während der Studie 190-201 hatten neun Patienten ein Grad-3-UE, aber kein Grad-4-UE. Bis zum Datenschnitt im Juni 2016 hatten 12 Patienten ein Grad-3-UE und ein Patient ein Grad-4-UE entwickelt (N=13 [54 %]). Ein Patient entwickelte einen Status epilepticus (Grad 4), dieser war aus Sicht des Prüfarztes allerdings nicht behandlungsbedingt.

Schwerwiegende unerwünschte Ereignisse

Während der Studie 190-201 traten bei 16 Patienten (67 %) insgesamt 32 SUEs auf. Bis zum Datenschnitt im Juni 2016 berichteten 19 der 24 Patienten (79 %) von insgesamt 51 SUEs. Überempfindlichkeitsreaktionen stellten die am häufigsten patientenrelevanten SUE mit einer Inzidenz von ≥ 10 % dar (Studie 190-201: 25 % und Studie 190-201/202: 29 %).

Vom pU beschriebene UE von besonderem Interesse

Die Daten beziehen sich primär auf den Datenschnitt vom 03. Juni 2016.

(1) Inzidenz von Überempfindlichkeits- und anaphylaktischen Reaktionen

Bei 14 der 24 Patienten (58 %) traten insgesamt 36 Überempfindlichkeitsreaktionen auf, diese waren allerdings nur bei drei Patienten schwer (CTCAE-Grad 3; N=3/24, 13 %).

(2) Inzidenz von UEs in zeitlichem Zusammenhang mit den Infusionen („temporally-related events", TRE)

Alle Patienten entwickelten mindestens ein TRE (N=24). 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.

(3) UEs in Zusammenhang mit dem ICV-Port

Insgesamt 11 Patienten (46 %) berichteten von insgesamt 30 UEs, die im Zusammenhang mit dem ICV-Port standen. Diese UEs waren bei 9 Patienten nur leichter Ausprägung (CTCAE-Grad 1: 38 %) und bei zwei Patienten (8 %) traten Grad-3-UEs (Infektionen) auf. 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.

(4) UEs von besonderem Interesse: Inzidenz von UEs, die das Nervensystem betreffen

Es traten keine Fälle von Hydrozephalus und Meningitis auf. Alle Patienten entwickelten mindestens ein UE innerhalb der SOC „Erkrankungen des Nervensystems", und diese waren mehrheitlich leicht bis mittelschwer (CTCAE-Grad 1 oder 2: n=20/24, 83 %). Bei einem der 24 Patienten (1 UE) wurde Status epilepticus als UE berichtet und bei 7 Patienten (29 %) insgesamt 13 Fälle von Petit-mal-Epilepsie. Im Rahmen der breiten MedDRA-Abfrage „Konvulsionen" (SMQ) wurden bis zum Datenschnitt 03. Juni 2016 insgesamt 362 UEs bei 23/24 Patienten (96 %) identifiziert. 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 %]). Sechs Patienten (25 %) entwickelten in der Studie 190-201 insgesamt 7 Konvulsionen, die nach der Implantation des ICV-Ports, aber vor der ersten Dosis Cerliponase alfa auftraten.

Zusammenfassende Einschätzung zur Sicherheit

Die Zulassung ist mit der Auflage zu weiteren Sicherheitsdaten verknüpft: Eine nicht-interventionelle post-authorisation Sicherheitsstudie (190-501) zur Beurteilung der Langzeitsicherheit einschließlich dem Auftreten von Überempfindlichkeitsreaktionen und Anaphylaxien auf Basis einer adäquaten Datenquelle, die sich aus einer Registrierung von Patienten mit CLN2 ergibt (Registerstudie). Ergebnisse sollen jährlich für eine Neubewertung vorgelegt werden. Zusätzlich eine post-authorisation Wirksamkeitsstudie (190-203) zur Beurteilung der Krankheitsprogression sowie der Sicherheit und Verträglichkeit, einschließlich mindestens N=5 Patienten unter 2 Jahren. Ergebnisse sind im Jahr 2020 zu erwarten [18].

Die Anzahl der mit Cerliponase alfa behandelten und untersuchten Patienten ist bisher gering. Informationen über Langzeitwirkungen liegen bis zum Datenschnitt 03. Juni 2016 mit einer medianen Behandlungsdauer von 95 Wochen vor. Es gab in der Studie keine Abbrüche aufgrund UE und keine Todesfälle. Bei 79 % der Patienten wurden insgesamt 51 SUEs und bei 54 % der Patienten wurden UEs des NCI-CTCAE-Grades 3 und höher erfasst. Eine Einschätzung über die Häufigkeit von insbesondere den UE von besonderem Interesse ist aufgrund der fehlenden Kontrollgruppe limitiert. Interessant ist, dass mit Amendement 4 der Studie 190-202 vom 17. März 2017 als ein weiteres UE vom besonderen Interesse kardiovaskuläre UEs und EKG-UEs hinzugefügt wurden.

Die EMA weist darauf hin, dass eine Kausalitätsbewertung der aufgetretenen UEs schwierig ist, da die Studie zum einen unkontrolliert ist und zum anderen unterschiedliche Ursachen für das Auftreten von UEs maßgeblich sein können. Es kann sowohl ein Zusammenhang des UEs mit dem Medikament bestehen, gleichzeitig aber auch mit dem ICV-Port, der Infusion, der zugrundeliegenden Erkrankung oder auch aufgrund von Begleitmedikationen. Im Summary of product characteristics (SmPC) wird explizit darauf verwiesen, dass vorsorgliche Maßnahmen getroffen werden sollen, um mögliche Überempfindlichkeitsreaktionen oder anaphylaktische Schocks zu vermeiden. Demnach wird empfohlen, dass die Gabe mit Antihistaminika mit oder ohne Antipyretika 30–60 Minuten vor jeder Infusion verabreicht werden sollte. Darüber hinaus verweist die EMA darauf, dass Cerliponase alfa bei Patienten mit Symptomen eines ICV-Lecks, nach Versagen oder bei ICV-bezogenen Infektionen kontraindiziert ist [18].

4 Anforderungen an eine qualitätsgesicherte Anwendung

Die Vorgaben der Fachinformation sind zu berücksichtigen.

Die Einleitung und Überwachung der Behandlung mit Cerliponase alfa darf nur durch Ärzte erfolgen, die Erfahrung in der intrazerebroventrikulären Anwendung von Arzneimitteln haben.

Dieses Arzneimittel wurde unter „besonderen Bedingungen“ zugelassen. Das bedeutet, dass weitere Nachweise für den Nutzen des Arzneimittels erwartet werden. Die Europäische Arzneimittel-Agentur wird neue Informationen zu diesem Arzneimittel mindestens jährlich bewerten und die Fachinformation, falls erforderlich, aktualisieren.

5 Zusammenfassung der Bewertung

Cerliponase alfa ist zugelassen für die Behandlung der neuronalen Ceroid-Lipofuszinose (NCL) Typ 2, auch als Tripeptidyl-Peptidase-1-(TPP1)-Mangel bezeichnet. Die Nutzenbewertung von Cerliponase alfa basiert auf der Zulassungsstudie 190-201, einer ‚first-in-human Phase 1/2-Studie' – einer multizentrischen, offenen, einarmigen Interventionsstudie, die sich in eine Dosis-Eskalationsphase und eine stabile Dosisphase gliedert und bereits abgeschlossen ist, sowie der sich daran anschließenden Extensionsstudie 190-202 mit einer Weiterführung der stabilen Dosis von 300 mg, die noch bis Dezember 2020 andauert. Eingeschlossen wurden Patienten zwischen drei und acht Jahren mit einer milden bis moderaten CLN2-Erkrankung unter der Annahme eines Scores von 3–6 auf der ML-Skala sowie einer bestätigten Genmutation des CLN2-/TPP1-Gens.

Zusätzlich wurde a priori im Originalprotokoll und SAP der Studie 190-201/190-202 festgelegt, dass für die primäre Auswertung ein Vergleich mit einer historischen Kontrollgruppe (190-201) vorgesehen war. Auf Empfehlung der Zulassungsbehörden wurde im ISE SAP für die primäre Auswertung eine 1:1-Paarbildung (Matching) der Teilnehmer der Studie 190-201/202 mit den Patienten der historischen Kontrollgruppe (Studie 190-901) festgelegt.

Die Zusammenfassung der Bewertung erfolgt differenziert für die einarmigen Studien 190-201/190-202 und für den gematchten Vergleich mit der historischen Kontrollgruppe aus Studie 190-901.

Ergebnisse der einarmigen Studien 190-201/190-202 (Datenschnitt: 03. Juni 2016)

Da es sich bei den Studien 190-201 und 190-202 um Studien ohne Kontrollgruppe handelt, wird von einem hohen Verzerrungspotential auf Studien- und Endpunktebene ausgegangen. Die Ergebnisse werden in Tabelle 31 zusammengefasst.

Tabelle 31: Zusammenfassende Darstellung der Studienergebnisse für die einarmigen Studien 190-201/190-202 (ITT-Population [N=23], Baseline: ab 300-mg-Dosierung)

Endpunkt	Studie 190-201[1)]	Studie 190-201/190-202[2)]
Mortalität		
Todesfälle[3)]	keine	keine
Morbidität		
Responderrate der ML-Skala (primärer Endpunkt, Responderanalyse)[4)]	Responder (bis Woche 48)[5)] n (%)	Responder (bis Woche 80)[6)] n (%)
	20 (87)	20 (87)
Abfall der ML-Skala (ursprünglicher primärer Endpunkt, Slopeanalyse)	skaliert auf 48 Wochen MW (SD)	skaliert auf 48 Wochen MW (SD)
	0,40 (0,81)	0,32 (0,46)
Veränderung der Punktzahl auf der ML -Skala	**Veränderung zu Baseline MW (SD)**	
	-0,4 (0,79): Woche 49 [N=23] -0,4 (0,9): Woche 73 [N=23] -0,5 (0,87): Woche 81 [N=17]	

Endpunkt	Studie 190-201[1)]	Studie 190-201/190-202[2)]
Lebensqualität		
	Veränderung von Studien -Baseline zu Woche 49 [N=23] MW (SD)	Veränderung von Studien-Baseline zu Woche 98 [N=12] MW (SD)
PedsQL (Parent Report for Toddlers)		
Körperliche Kompetenzen[7)]	-6,1 (21,7)	-27,9 (22,4)
Emotionale Kompetenzen[7)]	7,7 (26,2)	-0,5 (25,5)
Soziale Kompetenzen[7)]	7,8 (16,2)	-5,8 (18,3)
Schulische Kompetenzen[7)]	6,3 (29,8)[10)]	-17 (21,8)[10)]
Psychosozialer Summenscore[7), 8)]	7,9 (14,9)	-5,4 (11,1)
Gesamtscore[7) 9)]	2,6 (12,2)	-14,1 (13,1)
Sicherheit	**n (%)**	**n (%)**
Unerwünschte Ereignisse (UE)	24 (100)	24 (100)
UEs vom CTCAE-Grad ≥ 3	9 (38)	13 (54%)
Schwerwiegende UE (SUE)	16 (67)	19 (79)
UEs, die zum Studien- oder Behandlungsabbruch führen	0	0
UE, das zum Tod führte	0	0

[1)] Datenschnitt: 30. November 2015. Alle Patienten der Studie 190-201 (N=23, ITT) hatten zu diesem Datenschnitt die 48 Wochen der Behandlung mit 300 mg Cerliponase abgeschlossen.
[2)] Datenschnitt 03. Juni 2016, Behandlung über mindestens 72 Wochen mit 300 mg Cerliponase (ITT-Ppulation)
[3)] Todesfälle waren gemäß Studienprotokoll kein geplanter Effektivitätsparameter. Gleichwohl wurden Todesfälle als Sicherheitsendpunkt (UE) erfasst.
[4)] Responder waren definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 nicht im Netto um zwei oder mehr Punkte bzw. nicht auf einen Wert von Null abfiel (d. h. Response ist ein Nettoabfall um < 2 bis Woche 48, der während des Studienverlaufs einen Abfall von ≥ 2 Punkten beinhalten kann, gefolgt von einem Anstieg des Scores und einer resultierenden Gesamtabnahme von maximal 1 Punkt). Demgegenüber waren Non-Responder (bzw. stabiler Abfall) definiert als Anteil der Patienten, bei denen der ML-Score bis Woche 48 (Studientag 340 bezogen auf die erste 300-mg-Infusion) um zwei oder mehr Punkte (≥ 2) abfiel.
[5)] Die Responderanalyse bezieht sich auf den Studientag 340 bezogen auf die erste Infusion von 300 mg der Studie 190-201.
[6)] Die Responderanalyse bezieht sich auf den Studientag 540 bezogen auf die erste Infusion von 300 mg der Studie 190-201.
[7)] 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. Die zu erreichende Punktzahl liegt zwischen 0 und 100. Höhere Punktzahlen deuten auf eine bessere Lebensqualität hin.
[8)] Psychosozialer Summenscore setzt sich aus den Dimensionen Emotionale Kompetenzen, Soziale Kompetenzen und Schulische Kompetenzen zusammen.
[9)] Summe aller Dimensionen: Körperliche Kompetenzen, Emotionale Kompetenzen, Soziale Kompetenzen und Schulische Kompetenzen
[10)] Es fehlen sowohl zu Woche 49 und Woche 98 ein Patient: (Woche 49 [N=22]; Woche 98 [N=11])

Ergebnisse für den historischen Vergleich (190-201/190-202 vs. 190-901; Analyse vom 15. Juni 2016)

Die Ergebnisse für den historischen Vergleich (190-201/190-202 vs. 190-901) werden in der folgenden Tabelle unter Berücksichtigung des Verzerrungspotentials zusammengefasst. Es wird dargestellt, ob unter Cerliponase alfa ein statistisch signifikant positiver Effekt bei hohem oder unklarem (↑) oder niedrigem (↑↑) Verzerrungspotential bzw. ein statistisch signifikant negativer Effekt bei hohem oder unklarem (↓) oder niedrigem (↓↓) Verzerrungspotential bzw. kein Unterschied (↔) gezeigt werden konnte.

Tabelle 32: Zusammenfassende Darstellung der Studienergebnisse: Studie 190-201/190-202 (ITT) vs. 190-901 (auswertbare Patienten) nach 1:1-Paarbildung[3] (Dosis 300 mg[4], Datenschnitt 15. Juni 2016)

Endpunkt	Historischer Vergleich: Studie 190-201/190-202 [N=22] vs. Studie 190-901 [N=22]		Effekt
Mortalität	Keine Daten für einen Vergleich vorhanden		-
Morbidität			
Anteil der Responder (ML-Skala, primärer Endpunkt)[1]	Responder (Slope < 2 Punkte skaliert auf 48 Wochen), n (%)	Responder (Slope < 2 Punkte skaliert auf 48 Wochen), n (%)	
	22 (100)	10 (45)	↑
	RR: 2,14 [95%-KI 1,37; 3,36)][2] p-Wert: 0,0009[2]		
Abfall der ML-/HML-Skala (ursprünglicher primärer Endpunkt, Slopeanalyse)	skaliert auf 48 Wochen MW (SD) [95%-KI]	skaliert auf 48 Wochen MW (SD) [95%-KI][5]	
	0,34 (0,46) [0,13; 0,54]	2,0 (1,39) [1,38;2,62]	↑
	Differenz MW (SD): 1,66 (0,31) [95%-KI 1,02; 2,31][5] p-Wert: < 0,0001[6]		
Zeit bis zum stabilen Abfall um ≥ 2 Punkte oder Auftreten von Wert 0 auf ML-/HML-Skala (Beobachtungszeit bis 72 Wochen)	Mediane Zeit bis Ereignis: nicht erreicht	Mediane Zeit bis Ereignis: 285 Tage [95%-KI 210; 420]	
	HR: 0,10 [95%-KI 0,03; 0,38] p-Wert: 0,0005		↑
Veränderung der Punktzahl auf der ML-/HML-Skala	Veränderung zu Baseline, n (%) MW (SD)	Veränderung zu Baseline, n (%) MW (SD)	
	-0,5 (0,80): Woche 49 [N=22] -0,5 (0,91): Woche 73 [N=22] -0,4 (0,76): Woche 89 [N=17]	-2,1 (0,51): Woche 49 [N=21] -2,7 (0,67): Woche 73 [N=20] -2,7 (0,86): Woche 89 [N=11]	-
Lebensqualität	Keine Daten für einen Vergleich vorhanden		**-**
Sicherheit	Keine Daten für einen Vergleich vorhanden		-

[1] Geschätzte Neigung (Slope) kategorisiert nach < 2 und ≥ 2 Punkte auf der ML-/HML-Skala/48 Wochen
[2] Eigene Berechnungen. Korrektur für Zellen mit 0 Ereignissen durch Addition von 0,5.
[3] Matching-Methode #1 für Patienten aus der Cerliponase-Studie 190-201/202 und den historischen Kontrollen: Identischer ML-/HML-Wert und Altersunterschied so gering wie möglich, aber nicht mehr als 12 Monate, jeweils zu Baseline
[4] In der Studie 190-201/190-202 Baseline ab 300-mg-Dosierung; in der Studie 190-901 entsprach Baseline dem Alter (in Monaten) zum Zeitpunkt der 1:1-Paarbildung
[5] Zur Berechnung der KIs wurde eine t-Verteilung unter Einsatz der Satterthwaite-Approximation herangezogen.
[6] Zweistichproben-t-Test mit Satterthwaite-Approximation

Anhang

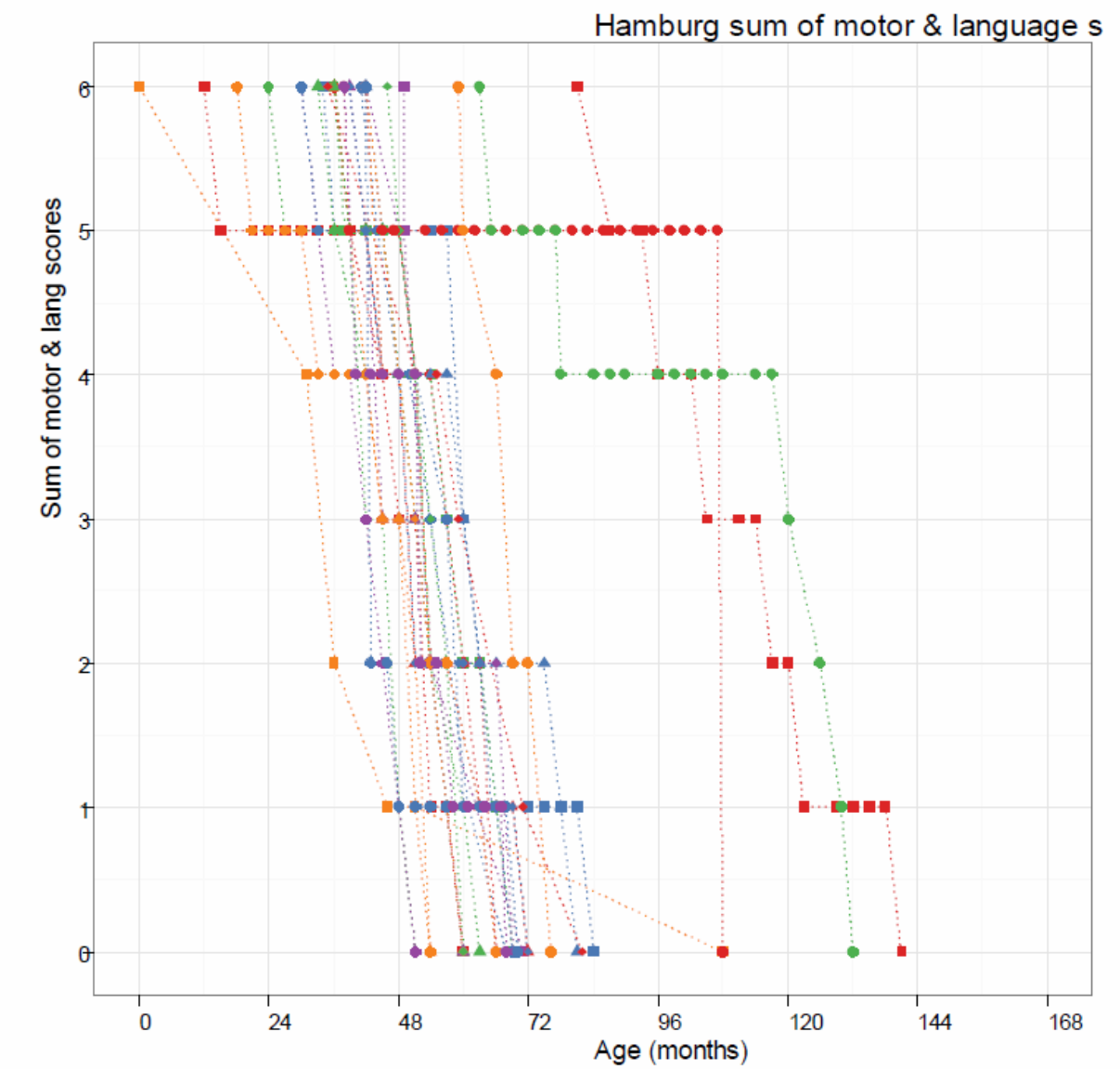


Abbildung 7: Individuelle Patientendaten der HML-Skala der 29 ausgewerteten Patienten der Studie 190-901 zum natürlichen Krankheitsverlauf

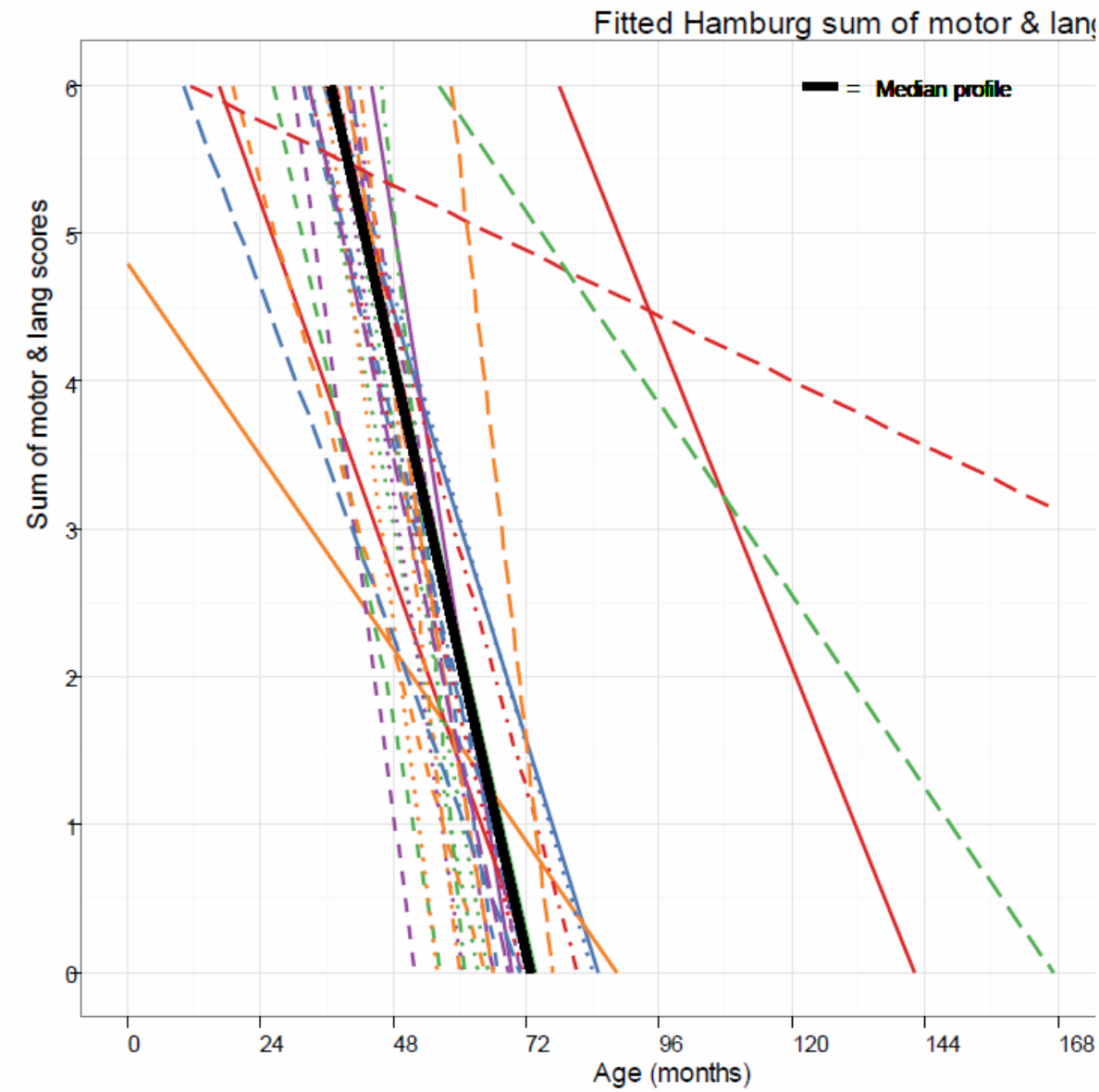


Abbildung 8: Mittels linearer Regression modellierte Verläufe der HML-Skala der 29 ausgewerteten individuellen Patientendaten der Studie 190-901 zum natürlichen Krankheitsverlauf

Referenzen

1. **BioMarin International.** Dossier zur Nutzenbewertung gemäß § 35a SGB V; Cerliponase alfa (Brineura); Modul 2; Allgemeine Angaben zum Arzneimittel, zugelassene Anwendungsgebiete. 26.06.2017.
2. **BioMarin International.** Dossier zur Nutzenbewertung gemäß § 35a SGB V; Cerliponase alfa (Brineura); 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. 26.06.2017.
3. **BioMarin Pharmaceutical.** BMN 190 Integrated summary of efficacy; final statistical analysis plan version 3. 25.04.2016.
4. **BioMarin Pharmaceutical.** Integrated summary of effectiveness. 30.11.2015.
5. **BioMarin Pharmaceutical.** Integrated summary of effectiveness. 15.06.2016.
6. **BioMarin Pharmaceutical.** A multicenter, multinational, extension study to evaluate the long-term efficacy and safety of BMN 190 in patients with CLN2 disease; clinical study protocol 190-202; amendment 3. 26.02.2016.
7. **BioMarin Pharmaceutical.** A multicenter, multinational, extension study to evaluate the long-term efficacy and safety of BMN 190 in patients with CLN2 disease; interim clinical study report 190-202. 09.05.2016.
8. **BioMarin Pharmaceutical.** A multicenter, multinational, extension study to evaluate the long-term efficacy and safety of BMN 190 in patients with CLN2 disease; study 190-202; addendum. 03.06.2016.
9. **BioMarin Pharmaceutical.** Natural history of late-infantile CLN2 disease: quantitative assesment of disease characteristics, rate of progression, and magnetic resonance imaging findings; natural history report 190-901. 30.06.2015.
10. **BioMarin Pharmaceutical.** 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 protocol 190-201; amendment 5. 24.09.2014.
11. **BioMarin Pharmaceutical.** 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. 27.04.2016.
12. **BioMarin Pharmaceutical.** 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; final statistical analysis plan 190-201; version 3.0. 27.04.2016.

13. **BioMarin Pharmaceutical.** Response to request for information from february 1, 2017 (to FDA). 2017.

14. **BioMarin Pharmaceutical.** Update analysis of the natural history of CLN2 disease: estimated rate of decline from the DEM-CHILD Multi-Center clinical NCL database; study 190-901; supplemental report. 01.05.2016.

15. **Chang M, Cooper JD, Davidson BL, van Diggelen OP, Elleder M, Goebel HH, et al.** CLN2. In: Mole S, Williams R, Goebel H (Eds). The neuronal ceroid lipofuscinoses (Batten Disease); second edition. Oxford University Press; 2011. S. 80-109.

16. **Connelly M, Rapoff MA.** Assessing health-related quality of life in children with recurrent headache: reliability and validity of the PedsQLTM 4.0 in a pediatric headache sample. J Pediatr Psychol 2006;31(7):698-702.

17. **European Medicines Agency (EMA).** Brineura: EPAR-product information EMEA/H/C/004065/0000 [online]. 22.06.2017. London (GBR): EMA. [Zugriff: 13.07.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/004065/WC500229798.pdf.

18. **European Medicines Agency (EMA).** Brineura: European public assessment report EMEA/H/C/004065/0000 [online]. 21.04.2017. London (GBR): EMA. [Zugriff: 13.07.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/004065/WC500229800.pdf.

19. **Haltia M, Goebel HH.** The neuronal ceroid-lipofuscinoses: a historical introduction. Biochimica et biophysica acta 2013;1832(11):1795-1800.

20. **Krohne TU, Herrmann P, Kopitz J, Rüther K, Holz FG.** Juvenile neuronale Zeroidlipofuszinose. Der Ophthalmologe 2010;107(7):606-611.

21. **Mapi Research Trust, Varni JW.** Scaling and scoring of the Pediatric Quality of Life Inventory PedsQL; Version 17 [online]. Lyon (FRA): Mapi Research Trust; 2017. [Zugriff: 29.08.2017]. URL: http://www.pedsql.org/PedsQL-Scoring.pdf.

22. **Preising MN, Lorenz B.** Genetik der neuronalen Zeroidlipofuszinosen. Der Ophthalmologe 2010;107(7):612-615.

23. **Schulz A, Kohlschütter A.** Neuronale Zeroidlipofuszinosen (NCL). Monatsschrift Kinderheilkunde 2012;160(8):734-741.

24. **Steinfeld R, Heim P, von Gregory H, Meyer K, Ullrich K, Goebel HH, et al.** Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations. American Journal of Medical Genetics 2002;112(4):347-354.

25. **Varni JW, Burwinkle TM, Seid M, Skarr D.** The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. Ambulatory Pediatrics: The Official Journal of the Ambulatory Pediatric Association 2003;3(6):329-341.

26. **Varni JW, Seid M, Kurtin PS.** PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Medical Care 2001;39(8):800-812.

27. **Varni JW, Seid M, Rode CA.** The PedsQL: measurement model for the pediatric quality of life inventory. Medical Care 1999;37(2):126-139.

28. **Williams RE, Adams HR, Blohm M, Cohen-Pfeffer JL, de Los Reyes E, Denecke J, et al.** Management strategies for CLN2 disease. Pediatr Neurol 2017;69:102-112.

29. **Worgall S, Kekatpure MV, Heier L, Ballon D, Dyke JP, Shungu D, et al.** Neurological deterioration in late infantile neuronal ceroid lipofuscinosis. Neurology 2007;69(6):521-535.

30. **Worgall S, Sondhi D, Hackett NR, Kosofsky B, Kekatpure MV, Neyzi N, et al.** Treatment of late infantile neuronal ceroid lipofuscinosis by CNS administration of a serotype 2 adeno-associated virus expressing CLN2 cDNA. Human Gene Therapy 2008;19(5):463-474.

31. **Wyrwich K, Evidera.** 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); Version 2.0 vom 16.05.2016.