

Nutzenbewertung

von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V

Bewertung von Arzneimitteln für seltene Leiden nach § 35a
Absatz 1 Satz 11 i. V. m. Kapitel 5 § 12 Nr. 1 Satz 2 Verfo

Wirkstoff: Atidarsagen autotemcel/OTL-200

Nutzenbewertung nach Fristablauf

Datum der Veröffentlichung: 1. Oktober 2026

Version 1.0

Inhaltsverzeichnis

Tabellenverzeichnis	4
Abbildungsverzeichnis	6
Abkürzungsverzeichnis	7
Hintergrund	9
1 Fragestellung	10
2 Auswahl und Methodik der relevanten Studien	11
2.1 Studienbasis für die Nutzenbewertung	11
2.2 Beschreibung der eingeschlossenen Studien	12
2.2.1 Integrierter Datensatz	12
2.2.2 Studie 201222	13
2.2.3 Studie CUP 207394	18
2.2.4 Studie CUP 206258	20
2.2.5 Studie HE 205029	23
2.2.6 Charakterisierung der Intervention im IDS	25
2.3 Endpunkte	27
2.3.1 Mortalität	28
2.3.2 Morbidität	28
2.3.3 Sicherheit	36
2.3.4 Erhebungszeitpunkte	39
2.4 Statistische Methoden	40
2.5 Verzerrungspotential auf Studien- und Endpunktebene	43
2.6 Indirekte Vergleiche	44
2.6.1 Geschwisterkinderanalyse IDS vs. TIGET-NHx (indirekter Vergleich C)	44
2.6.2 Zusammenfassende Bewertung der Methodik des indirekten Vergleichs	50
2.6.3 Frühsymptomatische Patientinnen und Patienten des IDS vs. TIGET-NHx (indirekter Vergleich B)	51
3 Ergebnisse der eingeschlossenen Studien	52
3.1 Studiencharakteristika und Studienmedikation	52
3.2 Mortalität	56
3.3 Morbidität	57
3.4 Lebensqualität	60
3.5 Sicherheit	60
4 Diskussion der Methodik und Ergebnisse	67
4.1 Design und Methodik der Studie	68
4.2 Zulassungsstatus und Zulassungspopulation	73
4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit	75
4.4 Mortalität	75
4.5 Morbidität	75
4.6 Lebensqualität	77
4.7 Sicherheit	77

5	Anforderung an eine qualitätsgesicherte Anwendung	80
6	Zusammenfassung der Nutzenbewertung	81
	Referenzen	83
	Anhang	87
	Begleitmedikation auf Einzelstudienebene	87
	GMFC-MLD.....	88
	Geschwisterkinderanalyse IDS vs. TIGET-NHx	89
	Studiencharakteristika und Studienmedikation	89
	Mortalität	92
	Morbidität	94
	Indirekter Vergleich fröhsymptomatische Patientinnen und Patienten	98

Tabellenverzeichnis

Tabelle 1:	Übersicht über die Studienbasis	11
Tabelle 2:	Charakterisierung der Studie 201222.....	13
Tabelle 3:	Für die Nutzenbewertung relevante Protokolländerungen der Studie 201222.....	16
Tabelle 4:	Charakterisierung der Studie CUP 207394.....	18
Tabelle 5:	Charakterisierung der Studie CUP 206258.....	20
Tabelle 6:	Für die Nutzenbewertung relevante Protokolländerungen der Studie CUP 206258.....	22
Tabelle 7:	Charakterisierung der Studie HE 205029	23
Tabelle 8:	Charakterisierung der Intervention im IDS (Studien 201222, CUP 207394, CUP 206258 und HE 205029).....	25
Tabelle 9:	Zusammenfassung der Endpunktbewertung der Studien 201222, CUP 207394, CUP 206258 und HE 205029.....	27
Tabelle 10:	Übersicht zu neuropsychologischen Untertests zur Indexbildung	34
Tabelle 11:	Erhebungszeitpunkte der berücksichtigten Endpunkte in den Studien 201222 und CUP 207394.....	40
Tabelle 12:	Erhebungszeitpunkte der berücksichtigten Endpunkte in den Studien CUP 206258 und HE 205029	40
Tabelle 13:	Allgemeine Angaben; IDS	52
Tabelle 14:	Charakterisierung der Studienpopulation; IDS, ITT-Population.....	53
Tabelle 15:	Angaben zur Exposition mit der Studienmedikation; IDS, ITT-Population	55
Tabelle 16:	Ergebnisse für „Gesamtüberleben“ im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); ITT-Population	57
Tabelle 17:	Ergebnisse für „GMFM“ (in %) im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); ITT-Population	57
Tabelle 18:	Ergebnisse der Nebenwirkungen – Zusammenfassung der UE im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); Sicherheitspopulation.....	60
Tabelle 19:	UE mit Inzidenz ≥ 20 % des gesamten IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); Sicherheitspopulation.....	61
Tabelle 20:	Schwere UE (Grad 3 oder höher), die bei ≥ 5 % der Teilpopulation und mindestens 2 Patientinnen und Patienten des gesamten IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029) aufgetreten sind, nach Systemorganklasse und Preferred Term im IDS; Sicherheitspopulation.....	63

Tabelle 21: SUE, die bei $\geq 5\%$ der Teilpopulation und mindestens 2 der Patientinnen und Patienten des gesamten IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029) aufgetreten sind, nach Systemorganklasse und Preferred Term im IDS; Sicherheitspopulation.....	65
Tabelle 22: Zusammenfassende Darstellung der Ergebnisse des IDS.....	81
Tabelle 23: Ergebnisse für „Anteil der Personen mit einem GMFC-MLD-Level ≥ 5 zum Zeitpunkt des letzten Follow-ups“ im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); ITT-Population.....	88
Tabelle 24: Allgemeine Angaben; Geschwisterkinderanalyse IDS vs. TIGET-NHx, MSAS.....	89
Tabelle 25: Charakterisierung der Studienpopulation; Geschwisterkinderanalyse IDS vs. TIGET-NHx, MSAS.....	89
Tabelle 26: Angaben zur Exposition mit der Studienmedikation; Geschwisterkinderanalyse IDS vs. TIGET-NHx, MSAS.....	92
Tabelle 27: Ergebnisse für „Alter zum Zeitpunkt des Todes“ zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258), 26.05.2025 (HE 205029) und 01.11.2021 (TIGET-NHx); MSAS.....	93
Tabelle 28: Ergebnisse für „Anteil der Personen mit einem GMFC-MLD-Level ≥ 5 zum letzten Follow-up“ und „Alter bis zum GMFC-MLD-Level ≥ 5 “ zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258), 26.05.2025 (HE 205029) und 01.11.2021 (TIGET-NHx); MSAS.....	94
Tabelle 29: Charakterisierung der Studienpopulation; indirekter Vergleich fröhsymptomatische Patientinnen und Patienten des IDS vs. TIGET-NHx, ITT.....	98

Abbildungsverzeichnis

Abbildung 1: Zusammensetzung des IDS [35]	13
Abbildung 2: Studienphasen der Studie 201222 [38]	16
Abbildung 3: Stufen des Klassifikationssystems zur Beschreibung der grobmotorischen Funktionen bei MLD [27]	29
Abbildung 4: Kaplan-Meier-Kurven zu „Alter zum Zeitpunkt des Todes“ im IDS und in der Studie TIGET-NHx; MSAS [27]	93
Abbildung 5: Kaplan-Meier-Kurven zum „Alter bis zum GMFC-MLD-Level ≥ 5 “ im IDS und in der Studie TIGET-NHx; MSAS [27].....	95

Abkürzungsverzeichnis

ALP	Alkalische Phosphatase
ALT	Alanin-Aminotransferase
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ARSA	Arylsulfatase A
AST	Aspartat-Aminotransferase
AUC	Area Under the Curve
BLA	Biologics License Application
BSC	Best Supportive Care
BSID	Bayley Scales of Infant and Toddler Development
CRF	Case Report Form (Fallberichtsbogen)
CSF	Zerebrospinalflüssigkeit
CSR	Clinical study report (Studienbericht)
CTCAE	Common Terminology Criteria for Adverse Events
CUP	Compassionate Use Program
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, Version 4
eCRF	electronic Case Report Form
EJ	Early Juvenile
EMA	European Medicines Agency
ENG	Elektroneurographie
EPAR	European Public Assessment Report
FDA	U.S. Food and Drug Administration
GBA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GMFC-MLD	Gross Motor Function Classification in MLD
GMFCS	Gross Motor Function Classification System
GMFM	Gross Motor Function Measure
HE	Hospital Exemption
HSPC	Hämatopoetische Stamm- und Vorläuferzellen
IDS	Integrierter Datensatz
INR	International Normalised Ratio
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
KI	Konfidenzintervall
LI	Late Infantile
LLOQ	Lower limit of quantification (untere Nachweisgrenze)
LS	Least squares
MAC	Myeloablative Konditionierung
MAS	Matched Analysis Set
MedDRA	Medical Dictionary for Regulatory Activities
MLD	Metachromatische Leukodystrophie
MSAS	Matched Sibling Analysis Set

NCV	Nerve Conduction Velocity
PBMC	Mononukleäre Zellen des peripheren Blutes
PT	Preferred Term
pU	pharmazeutischer Unternehmer
SAP	Statistischer Analyseplan
SD	Standardabweichung
SGB	Sozialgesetzbuch
SMAC	Sub-myeloablative Konditionierung
SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis
TIGET-NHx	Telethon Institute for Gene Therapy Natural History Study
UE	Unerwünschtes Ereignis
ULN	Upper Limit of Normal (Oberer Normalwert)
VEP	Visuell evozierte Potentiale
VerfO	Verfahrensordnung des GBA
WISC	Wechsler Intelligence Scale for Children
WPPSI	Wechsler Intelligence Preschool and Primary Scale of Intelligence

Hintergrund

Atidarsagen autotemcel/OTL-200 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 11 1. Halbs. SGB V gilt der medizinische Zusatznutzen durch die Zulassung als belegt.

Gemäß § 35a Absatz 2 SGB V entscheidet der Gemeinsame Bundesausschuss (GBA), 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 11 1. Halbs. SGB V, dass der Zusatznutzen eines Orphan Drug durch die Zulassung als belegt gilt, hat der GBA 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 GBA 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 GBA bewertet.

Bei Orphan Drugs erfolgt eine Beauftragung des IQWiG mit der Durchführung einer Nutzenbewertung bei zuvor festgelegter Vergleichstherapie erst dann, wenn der Umsatz des betreffenden Arzneimittels die Umsatzschwelle gemäß § 35a Absatz 1 Satz 12 SGB V überschritten hat und damit einer uneingeschränkten Nutzenbewertung unterliegt.

Der GBA bestimmt gemäß Kapitel 5 § 12 Absatz 1 Nummer 1 Satz 2 der Verfahrensordnung des GBA (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 Atidarsagen autotemcel/OTL-200 zieht der GBA 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 (pU) heran und bewertet die darin enthaltenen Daten nach Maßgabe der in Kapitel 5 § 5 Absatz 7 Nr. 1 bis 4 VerfO festgelegten Kriterien im Hinblick auf ihre therapeutische Relevanz.

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Atidarsagen autotemcel/OTL-200 in seiner Sitzung am 22. September 2026 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 30. Juni 2026 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pU in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

- Kosten der Therapie für die gesetzliche Krankenversicherung (GKV)
- Anzahl der GKV-Patientinnen und -Patienten in der Zielpopulation

Die Nutzenbewertung wird am 1. Oktober 2026 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des GBA (<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 GBA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1 Fragestellung

Atidarsagen autotemcel/OTL-200 (Libmeldy®), im Folgenden als OTL-200 bezeichnet, ist ein Arzneimittel mit dem Status „Orphan Drug“. Gemäß § 35a Absatz 1 Satz 11 1. Halbs. SGB V gilt für Orphan Drugs der medizinische Zusatznutzen bereits durch die Zulassung als belegt. Der GBA bestimmt bei Orphan Drugs, die einen Umsatz von 30 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. Die Bewertung bezieht sich auf folgende Indikation [5]:

Metachromatische Leukodystrophie (MLD), die durch Mutationen in beiden Allelen des Gens für die Arylsulfatase A (ARSA) gekennzeichnet ist, welche zu einer Verringerung der enzymatischen Aktivität von ARSA führt:

- a) bei Kindern mit im präsymptomatischen späten Säuglings- oder präsymptomatischen frühen Kindesalter auftretenden Formen,
- b) bei Kindern mit der im frühsymptomatischen frühen Kindesalter auftretenden Form, die jedoch noch selbstständig gehen können, vor dem Einsetzen einer kognitiven Verschlechterung.

Gemäß Fachinformation wird die zu infundierende Dosis auf Grundlage des Patientengewichts zum Zeitpunkt der Infusion festgelegt. Die empfohlene Mindestdosis von OTL-200 beträgt 3×10^6 CD34⁺-Zellen/kg und das maximal zu verabreichende Volumen von OTL-200 sollte stets < 20 % des geschätzten Plasmavolumens der Patientin oder des Patienten betragen.

Befristungsaufgaben

Für die erneute Nutzenbewertung nach Fristablauf sollte eine getrennte Ergebnisdarstellung der Zulassungsstudie 201222 (Teil des integrierten Datensatzes, IDS) für Patientinnen und Patienten mit und ohne klinische Manifestation der Erkrankung zu allen patientenrelevanten Endpunkten, die für den Nachweis eines Zusatznutzens herangezogen werden, durchgeführt werden. Insbesondere für Patientinnen und Patienten ohne klinische Manifestation der Erkrankung sollte eine Geschwisterkinderanalyse vorgelegt werden [8, 9, 11].

2 Auswahl und Methodik der relevanten Studien

2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studienname (Nummer)	Studie mit Dossier vom pU ein- gereicht	Studie vom pU als rele- vant für die Ableitung des Zusatz- nutzens erachtet	Studie relevant für die Nutzen- bewertung	Ausschlussgrund
Studien zum Wirkstoff				
201222 ¹⁾²⁾	Ja	Ja	Ja ³⁾	-
CUP 207394 ⁴⁾²⁾	Ja	Ja	Ja ³⁾	-
CUP 206258 ⁴⁾²⁾	Ja	Ja	Ja ³⁾	-
HE 205029 ⁴⁾²⁾	Ja	Ja	Ja ³⁾	-
205756 ⁴⁾	Ja	Nein ⁵⁾	Unklar	Hinweis auf ergebnisgesteuerte Berichterstattung: Finaler Datenschnitt vom 10.03.2026 nicht vorgelegt. ⁶⁾
OTL-200-10 ⁷⁾	Ja	Nein	Nein	Es liegen noch keine Ergebnisse der Langzeitnachbeobachtung vor. ⁸⁾
Studien zu externen Kontrollen				
TIGET-NHx ⁹⁾	Ja	Ja	Nein	Der indirekte Vergleich mit der Studie TIGET-NHx zum natürlichen Verlauf wird als nicht hinreichend valide betrachtet (Begründung siehe Kapitel 2.6).

¹⁾ Zulassungsrelevante Studie gemäß EPAR [4].

²⁾ Teil der gepoolten Auswertung in einem integrierten Datenset (IDS) bestehend aus den Studien 201222, CUP 207394, CUP 206258 und HE 205029.

³⁾ Die Ergebnisse der Studie werden im Rahmen der Auswertung im IDS berücksichtigt.

⁴⁾ Im Rahmen des Zulassungsverfahrens bei der EMA ergänzend zu der Zulassungsstudie berücksichtigt.

⁵⁾ In Modul 4 legt der pU als supportive Evidenz im Anhang eine kombinierte Geschwisterkinderanalyse des IDS einschließlich der Studie 205756 zur kryokonservierten Formulierung von OTL-200 vor (nachfolgend bezeichnet als Cryo-Studie). Diese Analyse erweitert die Geschwisterkinderanalyse des IDS um Daten von weiteren 10 präsymptomatischen, mit OTL-200 behandelten Patientinnen und Patienten mit LI- oder EJ-MLD mit Datenschnitt zum 22.08.2022.

⁶⁾ In Modul 4 gibt der pU für die Studie einen Datenschnitt vom 22.08.2022 an. Ein CSR liegt lediglich für einen Datenschnitt vom 06.06.2022 vor [40]. In diesem CSR wird das voraussichtliche Studienende mit dem Jahr 2028 angegeben, während ClinicalTrials.gov als tatsächliches Studienende den 10.03.2026 ausweist. Die Studie wurde gemäß pU vorzeitig beendet und der finale Datenschnitt vom 10.03.2026 entspricht der letzten Visite innerhalb der Cryo-Studie 205756. Gemäß pU war der Hintergrund für die Beendigung, dass zu diesem Zeitpunkt für die weitere Nachbeobachtung die Studie OTL-200-10 zur Verfügung stand, in welche Teilnehmende aus der Cryo-Studie übergehen konnten. Ergebnisse zum finalen Datenschnitt, der für die Nutzenbewertung relevant sein könnte, wurden jedoch nicht eingereicht. Darüber hinaus bleibt unklar, auf welchem Datenschnitt die in Modul 4 dargestellte Analyse „IDS + Cryo“ basiert. Aufgrund der fehlenden Nachvollziehbarkeit hinsichtlich der Datenauswahl und fehlenden Ergebnissen zum finalen Datenschnitt kann eine ergebnisgesteuerte Berichterstattung nicht ausgeschlossen werden. Die Ergebnisse der „Analyse IDS + Cryo“ werden daher vorliegend nicht für die Nutzenbewertung berücksichtigt.

⁷⁾ Die laufende prospektive Beobachtungsstudie LongTERM-MLD (OTL-200-10, EUPAS48374) wurde von der EMA im Oktober 2020 angefordert und startete im Dezember 2022. Sie dient der Beurteilung der langfristigen

klinischen Wirksamkeit und Sicherheit auf Basis eines Registers. Gemäß pU ist das geplante Ende der Studie für das Jahr 2042 prognostiziert, gemäß EMA für den 29. März 2041 [15].

⁸⁾ Der vorliegende CSR der LongTERM-MLD-Studie enthält eine integrierte Wirksamkeitsanalyse der Mutterstudien (201222, CUP 207394, CUP 206258, HE 205029 und 205756) mit einem Datenschnitt aus dem Jahr 2024 [34]. Diese Analyse umfasst jedoch ausschließlich Daten, die innerhalb der jeweiligen Nachbeobachtungszeiträume der Mutterstudien erhoben wurden (mindestens 8 Jahre). Somit liegen noch keine Daten vor, die im Rahmen der eigentlichen LongTERM-Studie erhoben wurden. Der Beginn der LongTERM-Studie hatte sich verzögert, um Rückmeldungen der FDA zum Studiendesign zum Zeitpunkt der Prüfung des Antrags auf Zulassung eines biologischen Arzneimittels abzuwarten. Im IQWiG-Bericht zum Neugeborenen-Screening wurde die integrierte Wirksamkeitsanalyse berücksichtigt, da für die Mutterstudien zu diesem Zeitpunkt lediglich Datenschnitte aus den Jahren 2022 bis 2024 vorlagen und die integrierte Analyse aus dem CSR der LongTERM-Studie somit für einige der Studien eine um etwa 1 bis 2 Jahre längere Beobachtungsdauer ermöglichte. Für die vorliegende Bewertung liegen hingegen für sämtliche Mutterstudien finale Datenschnitte aus dem Jahr 2025 vor. Die integrierte Analyse aus dem Jahr 2024 bietet daher gegenüber den verfügbaren finalen Datenschnitten keinen zusätzlichen Erkenntnisgewinn und wird nicht gesondert dargestellt.

⁹⁾ Der indirekte Vergleich der pivotalen Studie 201222 mit der Studie zum natürlichen Verlauf TIGET-NHx wurde innerhalb der Studienunterlagen der pivotalen Studie prädefiniert und geplant.

Abkürzungen: CSR: Clinical Study Report; EJ: Frühjuvenil (Early Juvenile); EMA: European Medicines Agency; EPAR: European Public Assessment Report; FDA: U.S. Food and Drug Administration; IDS: integrierter Datensatz; IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; MLD: Metachromatische Leukodystrophie; LI: Spätingfantil (Late Infantile); pU: pharmazeutischer Unternehmer

Zur Nutzenbewertung für OTL-200 herangezogene Unterlagen

- Dossier des pU zu OTL-200 [24, 25, 26, 27]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [3, 4]
- Studienbericht, inklusive Studienprotokoll und statistischem Analyseplan (SAP), der Studie 201222 [36, 37, 38, 39]
- Studienbericht, inklusive Studienprotokoll und SAP, der Studie CUP 207394 [29, 30, 31, 32, 33]
- Studienbericht, inklusive Studienprotokolle und SAP, der Studien CUP 206258 und HE 205029 [13, 14, 17, 18, 19, 20, 21, 22, 23, 41]
- SAP für den integrierten Datensatz [35]
- Fachinformation zu OTL-200 [5]

2.2 Beschreibung der eingeschlossenen Studien

Die Nutzenbewertung von OTL-200 im vorliegenden Anwendungsgebiet basiert auf einem integrierten Datensatz, der die vier Zulassungsstudien 201222, CUP 207394, CUP 206258 und HE 205029 umfasst. Die Studien und die Intervention werden in den Tabellen 2–8 charakterisiert.

2.2.1 Integrierter Datensatz

Analog zur vorherigen Nutzenbewertung zu OTL-200 [10, 28] legt der pU unter anderem gemeinsame Auswertungen für die Studienpopulationen der Studien 201222, CUP 207394, CUP 206258 und HE 205029 vor (siehe Abbildung 1). Alle Studien wurden im selben Behandlungszentrum durchgeführt. In Anbetracht einer Vergleichbarkeit der Studiendurchführung in eben jenem Zentrum und hinreichend ähnlicher Studienpopulationen erscheint die gemeinsame Auswertung der Studien grundsätzlich als adäquat.

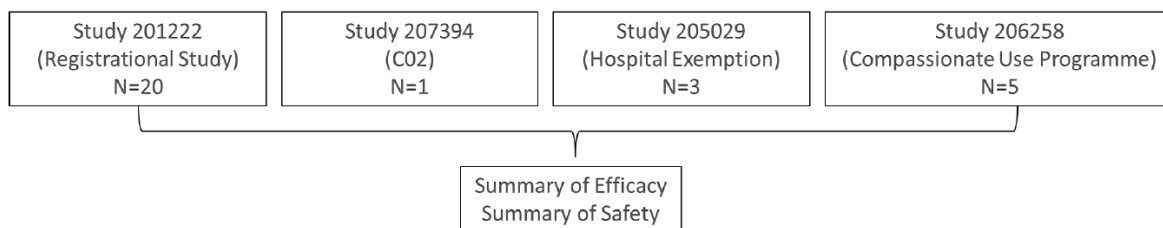


Abbildung 1: Zusammensetzung des IDS [35]

2.2.2 Studie 201222

Tabelle 2: Charakterisierung der Studie 201222

Charakteristikum	Beschreibung
Design und Studienablauf	Bei der Studie 201222 handelt es sich um eine nicht-randomisierte, offene, prospektive, einarmige, monozentrische Studie der Phase I/II mit Kindern mit LI- oder EJ-MLD zur Bestimmung der Wirksamkeit und Sicherheit von OTL-200. Als Vergleichsarm der Studie 201222 wurde prospektiv die natürliche Verlaufskohorte TIGET-NHx (siehe Kapitel 2.6) geplant. Die Studie unterteilt sich in die folgenden, zum Teil überlappenden Phasen: • Screeningphase (unbekannter Zeitraum): Überprüfung der Ein-/ Ausschlusskriterien • Baselinephase (Ende der Screeningphase bis Tag -5): Erhebung von Baseline-Messwerten • Behandlungsphase (Tag -35 bis Tag 0): ◦ Knochenmarkentnahme (einschließlich Ersatzprobe mit Ersatzstammzellen) ◦ Isolierung von CD34+-Zellen aus einer Knochenmarkprobe ◦ Konditionierung mit Busulfan ◦ Verabreichung der OTL-200-Infusion • Follow-up-Phase: Nachbeobachtung nach OTL-200-Infusion für mindestens 8 Jahre¹⁾ Der Studienverlauf ist zudem in Abbildung 2 dargestellt. Es wurden die folgenden Kriterien für den Abbruch der gesamten Studie festgelegt: • 2 transplantationsbezogene Todesfälle pro 4 behandelte Patientinnen und Patienten. • a) Ein Fall maligner Proliferation pro 2 behandelte Patientinnen und Patienten, die mit dem Gentransfer im Zusammenhang stehend erachtet werden, oder b) 2 Fälle maligner Proliferation, die mit der Gentherapie im Zusammenhang stehend erachtet werden. Die Studie ist abgeschlossen.
Population	Wesentliche Einschlusskriterien • Präsymptomatische Patientinnen und Patienten mit LI-MLD. • Prä- oder frühsymptomatische Patientinnen und Patienten mit EJ-MLD. • MLD-Diagnose basierend auf ◦ der Bestimmung der ARSA-Enzymaktivität unterhalb des Normalbereichs und ◦ der Identifizierung von 2 krankheitsverursachenden Mutationen. • LI-Form, definiert durch Vorliegen von 2 der 3 folgenden Kriterien: ◦ Alter bei Auftreten von Symptomen beim älteren Geschwisterkind ≤ 30 Monate und/oder ◦ 2 Null (0)-mutierte ARSA-Allele und/oder ◦ periphere Neuropathie laut ENG.

Charakteristikum	Beschreibung										
	- EJ-Form, definiert durch das Vorliegen von 2 der 3 folgenden Kriterien: - Alter bei Auftreten von Symptomen bei der Testperson oder beim älteren Geschwisterkind zwischen 30 Monaten und 6 Jahren und/oder 1 Null (0)- und 1 R-mutiertes ARSA-Allel und/oder - periphere Neuropathie laut ENG. - Präsymptomatische LI- oder EJ-Form, definiert als fehlende neurologische Beeinträchtigung (krankheitsbedingte Symptome), mit oder ohne Anzeichen der Erkrankung, die durch ENG- und MRT-Untersuchungen nachgewiesen wurden. - Kinder mit fröhsymptomatischer EJ wurden nur berücksichtigt, wenn sie - noch einen IQ ≥ 70 aufwiesen und - noch in der Lage waren, unabhängig ≥ 10 Schritte gehen zu können. Wesentliche Ausschlusskriterien - Positiver Befund für HIV, Hepatitis-C- und/oder Hepatitis-B-Virus. - Neoplastische Erkrankung. - Typische für das myelodysplastische Syndrom oder für die akute myeloische Leukämie zytogenetische Veränderungen. - Endorgandysfunktionen oder andere schwere Krankheiten, die eine Studienteilnahme nach Einschätzung der klinischen Prüfperson nicht gestattet. - Beteiligung an anderen Studien oder anderen therapeutischen Ansätzen, die verfügbar werden könnten. - Allogene hämatopoetische Stammzelltransplantation in den letzten 6 Monaten. - Allogene hämatopoetische Stammzelltransplantation mit Nachweis von residualen allogenen Zellen.										
Intervention und Zahl der Patientinnen und Patienten	Gescreent gesamt: N = 22 Eingeschlossen: N = 22 Intervention: N = 20 ²⁾										
Ort und Zeitraum der Durchführung; Datenschnitte	Studienzentren 1 Studienzentrum in Mailand, Italien Studienzeitraum - Einschluss erste Testperson: 09.04.2010 - Einschluss letzte Testperson: keine Angabe (spätestens Mai 2016)³⁾ - End of Study⁴⁾ (Database lock): 19.09.2025 Datenschnitte: <table> <tr> <td>Interim-Datenschnitt 1</td><td>24.03.2017 (prädefiniert)⁵⁾</td></tr> <tr> <td>Interim-Datenschnitt 2</td><td>30.03.2018⁶⁾ (GBA-Erstbewertung)⁷⁾</td></tr> <tr> <td>Interim-Datenschnitt 3</td><td>04.03.2022⁸⁾</td></tr> <tr> <td>Interim-Datenschnitt 4</td><td>29.03.2024⁹⁾</td></tr> <tr> <td>Finaler Datenschnitt</td><td>30.06.2025¹⁰⁾</td></tr> </table>	Interim-Datenschnitt 1	24.03.2017 (prädefiniert) ⁵⁾	Interim-Datenschnitt 2	30.03.2018 ⁶⁾ (GBA-Erstbewertung) ⁷⁾	Interim-Datenschnitt 3	04.03.2022 ⁸⁾	Interim-Datenschnitt 4	29.03.2024 ⁹⁾	Finaler Datenschnitt	30.06.2025 ¹⁰⁾
Interim-Datenschnitt 1	24.03.2017 (prädefiniert) ⁵⁾										
Interim-Datenschnitt 2	30.03.2018 ⁶⁾ (GBA-Erstbewertung) ⁷⁾										
Interim-Datenschnitt 3	04.03.2022 ⁸⁾										
Interim-Datenschnitt 4	29.03.2024 ⁹⁾										
Finaler Datenschnitt	30.06.2025 ¹⁰⁾										
Endpunkte gemäß Studienprotokoll	Primäre Endpunkte - Verbesserung im GMFM-Wert um 10 % im Vergleich mit der historischen Kontrolle 24 Monate nach der Behandlung. - Anstieg der ARSA-Restaktivität im Vergleich zu den Werten vor der Behandlung um ≥ 2 SD 24 Monate nach der Behandlung. - Unerwünschte Ereignisse und weitere Sicherheitsparameter										

Charakteristikum	Beschreibung
	Sekundäre Endpunkte • Nervenleitgeschwindigkeits-Index im Vergleich zur historischen Kontrolle 24 Monate nach der Behandlung. • MRT-Gesamtwert im Vergleich zur historischen Kontrolle 24 Monate nach der Behandlung. • GMFC-MLD-Werte in verschiedenen Altersstufen im Vergleich zur historischen Kontrolle. • Messung eines IQ über 55 bei neuropsychologischen Tests nach 24, 30 und 36 Monaten. • Engraftment der autologen Vorläuferzellen, die aus dem Knochenmark entnommen und mit einem lentiviralen Vektor transduziert wurden, um mehr als 4 % 12 Monate nach der Transplantation. • Korrelationen zwischen dem Ausmaß des Engraftments transduzierter Zellen und der Busulfan-Exposition. Im SAP zusätzlich beschrieben: • Alter zum Zeitpunkt des Todes im Vergleich zur historischen Kontrolle. Explorative Endpunkte • Elektroenzephalographie • Visuell evozierte Potentiale • Akustisch evozierte Hirnstammpotentiale

¹⁾ Mit Protokollversion 10 (07.09.2017) wurde etwa 7 ½ Jahre nach Einschluss der ersten Studienperson eine Langzeit-Nachbeobachtungsdauer von bis zu 15 Jahren eingeführt. Nach Studienjahr 8 sollten die Teilnehmenden einmal jährlich per Brief, E-Mail, Telefon oder auf anderem verfügbaren Wege kontaktiert und zu ausgewählten Endpunkten befragt werden. Zudem konnten behandelte Patientinnen und Patienten gemäß Studienprotokoll in eine separate Langzeit-Nachbeobachtungsstudie (Studie OTL-200-10, Studienstart im Dezember 2022) übergehen, sofern sie mindestens 8 Jahre in der Studie 201222 absolviert hatten.

²⁾ 2 der 22 Kinder wurden vor Behandlung aus der Studie ausgeschlossen (siehe Kapitel 3.1).

³⁾ Gemäß Protokollversion 8 (03.10.2014) ist das Rekrutierungsende auf den Mai 2016 datiert.

⁴⁾ Das Studienende (end of study) wurde definiert als das Datum der Datenbanksperre nach der letzten Visite der letzten Studienperson.

⁵⁾ Prädefinierte Interim-Analyse, nachdem mindestens 8 Patientinnen und Patienten eine Nachbeobachtungszeit von mindestens 2 Jahren absolviert hatten, um die Wirksamkeit und Sicherheit des Prüfpräparats zu bewerten. Mit Protokollversion 8 (03.10.2014) wurde die geplante Interim-Analyse verlegt, sodass anstelle einer 3-jährigen Nachbeobachtungszeit nun eine 2-jährige Nachbeobachtung der ersten 8 rekrutierten Studienteilnehmenden als Kriterium galt.

⁶⁾ Alle (überlebenden) Patientinnen und Patienten hatten eine Nachbeobachtungszeit von mindestens 3 Jahren absolviert; Anlass war ein Antrag auf Marktzulassung.

⁷⁾ Datenschnitt wurde in der vorangegangenen Nutzenbewertung zu OTL-200 berücksichtigt [10].

⁸⁾ Anlass war ein Antrag auf Zulassung eines biologischen Arzneimittels (Biologics License Application, BLA) bei der FDA. Zudem wurden Daten für die integrierten Wirksamkeits- und Sicherheitsanalysen im Rahmen des BLA mit einem Datenschnitt am 01. November 2022 extrahiert. Für den 120-Tage-Sicherheitsbericht zum BLA war der Datenschnitt der 15. Mai 2023.

⁹⁾ Anlass war ein Antrag auf Verlängerung der EU-Zulassung.

¹⁰⁾ Datenschnitt gemäß Modul 4. Dem CSR lässt sich ein konkreter Datenschnitt nicht entnehmen; es ist lediglich ein Datum für die Datenbanksperre angegeben (19.09.2025).

Abkürzungen: ARSA: Arylsulfatase A; BLA: Biologics License Application; CSR: Clinical Study Report; EJ: Early Juvenile; ENG: Elektroneurographie; FDA: U.S. Food and Drug Administration; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; LI: Late Infantile; MLD: Metachromatische Leukodystrophie; MRT: Magnetresonanztomographie; SAP: Statistischer Analyseplan; SD: Standardabweichung.

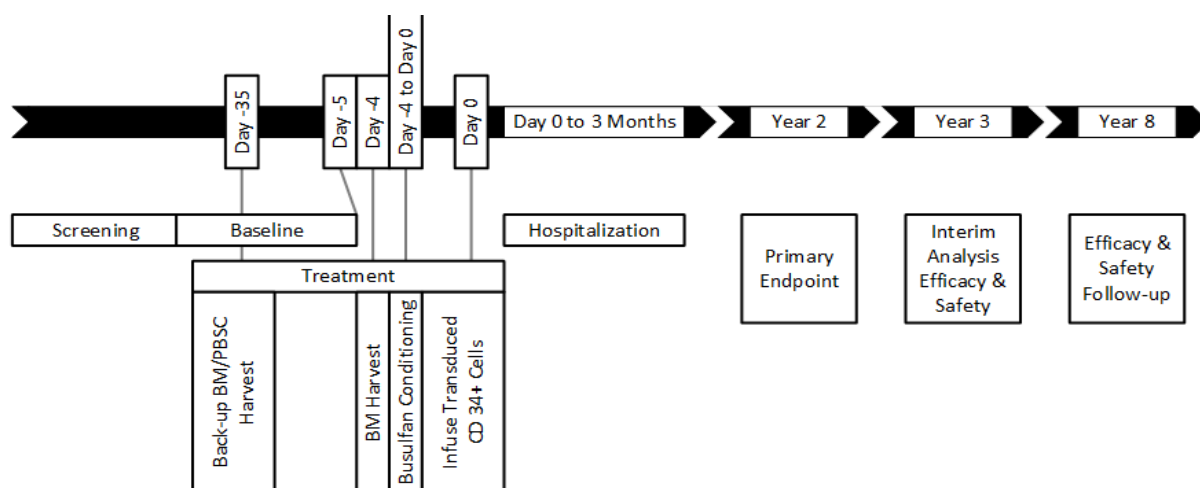


Abbildung 2: Studienphasen der Studie 201222 [38]

Protokolländerungen

Das erste Studienprotokoll ist auf den 18. Oktober 2009 datiert. Noch vor Beginn der Studie (09. April 2010) wurde ein erstes Protokoll-Amendment am 26. Januar 2010 erstellt. Diese Protokollversion 2 ist gemäß Studienbericht als Originalstudienprotokoll zu betrachten. Nach Studienbeginn (also nach Einschluss der ersten Studienperson am 09. April 2010) erfolgten 11 weitere Protokoll-Amendments mit einer Änderung der Hauptversionsnummer. Die aktuelle Protokollfassung liegt in Version 13.3 vor. Wesentliche Protokolländerungen sind in Tabelle 3 dargestellt.

Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie 201222

Amendment	Wesentliche Änderungen
Version 3 vom 04.04.2012 (Anzahl bis dahin eingeschlossene Personen: 5)	- Einführung des Instruments GMFC-MLD und der neuropsychologischen Tests BSID, WPPSI, WISC-III.¹⁾ - Nutzung der zweiten Stammzellentnahme („Back-up“) nicht nur im Falle einer Kontamination, sondern auch sofern die erste Entnahme nicht ausreichend für die initiale Gesamtzahl der CD34+-Zellen war.²⁾ - Einführung GMFM-Erhebung zu Monat 3 und zu Monat 9.³⁾
Version 4 vom 10.01.2013 (Anzahl bis dahin eingeschlossene Personen: 8)	Keine relevanten Änderungen.
Version 5 vom 10.03.2013 (Anzahl bis dahin eingeschlossene Personen: 10)	- Verlängerung der Nachbeobachtungszeit von 3 auf 8 Jahre.⁴⁾ - Einführung einer Interim-Analyse nach 3 Jahren Nachbeobachtung.
Version 6 vom 10.12.2013 (Anzahl bis dahin eingeschlossene Personen: 13)	- Einführung der Anforderung an die klinische Prüfperson, die Eignung von frühsymptomatischen EJ-Patientinnen und -Patienten unmittelbar vor der Behandlung erneut zu überprüfen und Personen aus der Studie (vor der Behandlung) auszuschließen, wenn es zwischen Studieneinschluss und geplanter Behandlung zu einem schnellen und schweren Fortschreiten der Erkrankung kam.⁵⁾ - Entfernung der Bedingung, dass frühsymptomatische Symptome nicht länger als 6 Monate bestehen durften, und Einführung von zusätzlichen Kriterien, die erfüllt sein mussten,

Amendment	Wesentliche Änderungen
	damit frühsymptomatische Kinder noch für einen Studieneinschluss berücksichtigt wurden⁵⁾: ○ IQ $\geq$ 70 Punkte und ○ Fähigkeit, unabhängig $\geq$ 10 Schritte gehen zu können. • Einführung, dass 2 von 3 Kriterien für die Diagnosestellung einer LI- oder EJ-MLD erfüllt sein mussten⁶⁾: ○ Alter des älteren Geschwisterkinds bei Symptom-Auftreten ○ Genotyp ○ Vorhandensein einer peripheren Neuropathie • Umstellung des Konditionierungsschemas von SMAC auf MAC.⁷⁾ • Einführung von Erhebungen des IQ nach 2, 2,5 und 3 Jahren.⁸⁾
Version 7 vom 10.01.2014 (Anzahl bis dahin eingeschlossene Personen: 14)	Keine relevanten Änderungen.
Version 8 vom 03.10.2014 (Anzahl bis dahin eingeschlossene Personen: 21)	Präzisierung der Kriterien für frühsymptomatische EJ-Patientinnen und -Patienten und Überprüfung der Eignung von LI-Patientinnen und -Patienten hinsichtlich einer bedeutsamen Krankheitsprogression zwischen Screening und Baseline. ⁹⁾
Version 9 vom 08.02.2016 (Anzahl bis dahin eingeschlossene Personen: k. A.) ¹⁰⁾	Keine relevanten Änderungen.
Version 10 vom 07.09.2017 (Anzahl bis dahin eingeschlossene Personen: 22) ¹⁰⁾	Keine relevanten Änderungen.
Version 11 vom 02.06.2018 (Anzahl bis dahin eingeschlossene Personen: 22) ¹⁰⁾	Keine relevanten Änderungen.
Version 12 vom 20.04.2020 (Anzahl bis dahin eingeschlossene Personen: 22) ¹⁰⁾	Keine relevanten Änderungen.
Version 13 vom 11.11.2020 (Anzahl bis dahin eingeschlossene Personen: 22) ¹⁰⁾	Keine relevanten Änderungen.

¹⁾ Es ist unklar, ob und wie die Baseline-Daten bisher eingeschlossener Studienteilnehmenden erfasst wurden.

²⁾ Es ist unklar, wie vor dem Amendment zu Protokollversion 3 mit einer unzureichenden Stammzellentnahme umgegangen wurde.

³⁾ Es ist unklar, wie viele Personen bereits $\geq$ 3 Monate in der Studie waren und fehlende Erhebungen zu diesen Messzeitpunkten hatten.

⁴⁾ Mit Studienbeginn am 09.04.2010 sollte die Verlängerung rechtzeitig vor Ende der 3-jährigen Nachbeobachtungszeit der ersten Studienteilnehmenden stattgefunden haben, sodass alle bis dahin eingeschlossenen Personen von der verlängerten Nachbeobachtung umfasst sein sollten.

⁵⁾ Es ist unklar, ob die Kinder, die vor dem Amendment zu Protokollversion 6 in die Studie aufgenommen wurden, diese Kriterien erfüllten oder sich in ihrer Symptomatik wesentlich von Kindern unterschieden, die nach Protokollversion 6 rekrutiert wurden. Laut Modul 4 waren bis zu diesem Amendment insgesamt 2 Kinder gemäß der vorherigen Definition als frühsymptomatisch eingeschlossen worden.

⁶⁾ Es ist unklar, ob Kinder vor dem Amendment zu Protokollversion 6 alle 3 Kriterien erfüllen mussten und ob diese Änderung damit zu einer relevanten Veränderung der Studienpopulation führte.

⁷⁾ Zur Beurteilung von Effekten legte der pU Subgruppenanalysen vor.

⁸⁾ Es ist unklar, wie viele Personen bereits ≥ 2 Jahre in der Studie waren und fehlende Erhebungen zu diesen Messzeitpunkten hatten.

⁹⁾ Klarstellung, dass sich die Definition für frühsymptomatisch ausschließlich auf EJ-Patientinnen und -Patienten bezieht.

¹⁰⁾ Das Rekrutierungsende war laut Amendment zu Protokollversion 8 auf Mai 2016 datiert.

Abkürzungen: BSID: Bayley Scales of Infant and Toddler Development; EJ: Early Juvenile; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; k. A.: keine Angabe; LI: Late Infantile; MAC: Myeloablative Konditionierung; MLD: Metachromatische Leukodystrophie; pU: pharmazeutischer Unternehmer; SMAC: Sub-myeloablative Konditionierung; WISC: Wechsler Intelligence Scale for Children; WPPSI: Wechsler Intelligence Preschool and Primary Scale of Intelligence.

2.2.3 Studie CUP 207394

Tabelle 4: Charakterisierung der Studie CUP 207394

Charakteristikum	Beschreibung
Design und Studienablauf	Bei der Studie CUP 207394 handelt es sich um den Fallbericht eines im Rahmen eines Härtefallprogramms mit OTL-200 behandelten Kindes mit frühsymptomatischer EJ-MLD. Das Härtefallprogramm (CUP) wurde ermöglicht, da die Rekrutierung von EJ-Patientinnen und -Patienten für die Studie 201222 zu diesem Zeitpunkt beendet war.¹⁾ Die Studie unterteilt sich in Anlehnung an die Studie 201222 in die folgenden, zum Teil überlappenden Phasen: • Screeningphase (unklarer Zeitraum): Überprüfung der Patienteneignung • Baselinephase (nach der Screeningphase bis Tag -5): Erhebung von Baseline-Messwerten • Behandlungsphase (Tag -35 bis Tag 1): ◦ Mobilisation des peripheren Blutes mit anschließender Apherese zur Gewinnung von Ersatzstammzellen für Ersatzprobe ◦ Knochenmarkentnahme für Behandlungsprobe ◦ Isolierung von CD34+-Zellen aus einer Knochenmarkprobe ◦ Konditionierung mit Busulfan (SMAC) ◦ Verabreichung der OTL-200-Infusion • Follow-up-Phase: Nachbeobachtung nach der OTL-200-Infusion für mindestens 8 Jahre²⁾ Die Studie ist abgeschlossen.
Population	Es sind keine formalen Ein- und Ausschlussgründe vorhanden. Es wurden die folgenden Parameter vor Studieneinschluss überprüft: • Allgemeine Untersuchung (Anamnese, körperliche Untersuchung und Lansky-Performance-Status) • Laboruntersuchungen • Knochenmarksuntersuchung • Peripherer zytogenetischer Bluttest • Mikrobiologische Untersuchung • Krankheitsspezifische Untersuchungen • Sequenzierung der ARSA-codierenden Region ◦ Neurologische Untersuchung ◦ ENG

Charakteristikum	Beschreibung								
	Laut Studienbericht erfüllte der Patient die Ein- und Ausschlusskriterien der Studie 201222 mit folgender Ausnahme: Zum Zeitpunkt des Studieneinschlusses wies der Patient bereits Symptome über einen Zeitraum von 8 Monaten auf; dies überstieg die zum damaligen Zeitpunkt festgelegte Grenze von 6 Monaten für das Vorliegen von Symptomen. ³⁾								
Intervention und Zahl der Patientinnen und Patienten	Eingeschlossene Testpersonen: N = 1 Personen, die OTL-200 erhielten: N = 1								
Ort und Zeitraum der Durchführung; Datenschnitte	Studienzentren 1 Studienzentrum in Mailand, Italien Studienzeitraum • Erster Patient erste Visite: 23.04.2013 • Letzter Patient letzte Visite 19.06.2025 Datenschnitte <table> <tr> <td>Interim-Datenschnitt 1</td><td>05.01.2018⁴⁾ (GBA-Erstbewertung)⁵⁾</td></tr> <tr> <td>Interim-Datenschnitt 2</td><td>k. A. (Datum Interim-CSR: 10.03.2023)⁶⁾</td></tr> <tr> <td>Interim-Datenschnitt 3</td><td>29.03.2024⁷⁾</td></tr> <tr> <td>Finaler Datenschnitt</td><td>30.06.2025⁸⁾</td></tr> </table> Der vorliegende Studienbericht ist auf den 27.01.2026 datiert.	Interim-Datenschnitt 1	05.01.2018 ⁴⁾ (GBA-Erstbewertung) ⁵⁾	Interim-Datenschnitt 2	k. A. (Datum Interim-CSR: 10.03.2023) ⁶⁾	Interim-Datenschnitt 3	29.03.2024 ⁷⁾	Finaler Datenschnitt	30.06.2025 ⁸⁾
Interim-Datenschnitt 1	05.01.2018 ⁴⁾ (GBA-Erstbewertung) ⁵⁾								
Interim-Datenschnitt 2	k. A. (Datum Interim-CSR: 10.03.2023) ⁶⁾								
Interim-Datenschnitt 3	29.03.2024 ⁷⁾								
Finaler Datenschnitt	30.06.2025 ⁸⁾								
Endpunkte gemäß Studienprotokoll	Es wurden keine Endpunkte für die Studie CUP 207394 bestimmt. Laut Studienbericht wurden, sofern angebracht, Endpunkte der Studie 201222 herangezogen.								

¹⁾ Zu diesem Zeitpunkt konnten in die Studie 201222 nur LI-Patientinnen und -Patienten eingeschlossen werden.

²⁾ Mit Protokollversion 2 (07.10.2016) wurde die Nachbeobachtungszeit von 12 Monaten auf 8 Jahre erhöht und mit Protokollversion 3 (12.02.2018) wurde knapp 5 Jahre nach Einschluss der Studienperson (23.04.2013) die Langzeit-Nachbeobachtungsdauer von „bis zu“ 8 auf „mindestens“ 8 Jahre verlängert. Mit Studienstart der Langzeit-Nachbeobachtungsstudie OTL-200-10 im Dezember 2022 hatte die Studienperson zudem die Möglichkeit, in diese überzugehen.

³⁾ Dieses Einschlusskriterium wurde mit Protokoll-Amendment 6 der Studie 201222 entfernt (vgl. Tabelle 3).

⁴⁾ Anlass war ein Antrag auf Marktzulassung.

⁵⁾ Datenschnitt wurde in der vorangegangenen Nutzenbewertung zu OTL-200 berücksichtigt [10].

⁶⁾ Anlass war ein Antrag auf Zulassung eines biologischen Arzneimittels (Biologics License Application, BLA) bei der US-amerikanischen FDA. Zudem wurden Daten für die integrierten Wirksamkeits- und Sicherheitsanalysen im Rahmen des BLA mit einem Datenschnitt am 01. November 2022 extrahiert. Für den 120-Tage-Sicherheitsbericht zum BLA war der Datenschnitt der 15. Mai 2023.

⁷⁾ Anlass war ein Antrag auf Verlängerung der EU-Zulassung.

⁸⁾ Datenschnitt gemäß Modul 4. Dem CSR lässt sich ein konkreter Datenschnitt nicht entnehmen; es ist lediglich ein Datum für die Datenbanksperre angegeben (19.09.2025).

Abkürzungen: ARSA: Arylsulfatase A; BLA: Biologics License Application; CSR: Clinical Study Report; EJ: Early Juvenile; ENG: Elektroneurographie; FDA: U.S. Food and Drug Administration; k. A.: keine Angabe; LI: Late Infantile; MLD: Metachromatische Leukodystrophie; SMAC: Sub-myeloablative Konditionierung.

Protokolländerungen

Das Studienprotokoll liegt in den Versionen 1, 2 und 3 vor. Im vorliegenden Studienbericht (27. Januar 2026) sind Änderungen gegenüber dem initialen Studienprotokoll (ohne Datum) zusammengefasst dargestellt. In Folge von Amendment 1 (07. Oktober 2016) und Amendment 2 (12. Februar 2018) wurde die Nachbeobachtungszeit sukzessive von 12 Monaten auf 8 Jahre und schließlich auf mindestens 8 Jahre verlängert.

2.2.4 Studie CUP 206258

Tabelle 5: Charakterisierung der Studie CUP 206258

Charakteristikum	Beschreibung
Design und Studienablauf	Bei der Studie CUP 206258 handelt es sich um eine Fallserie von präsymptomatischen Erkrankten mit früheinsetzender LI- und EJ-MLD, die im Rahmen eines Härtefall-programms (CUP) mit OTL-200 behandelt wurden. Die Studie unterteilt sich in Anlehnung an die Studie 201222 in die folgenden, zum Teil überlappenden Phasen: • Screeningphase (unklarer Zeitraum): Überprüfung der Patienteneignung • Baselinephase (Ende des Screenings bis Tag -5): Erhebung von Baseline-Messwerten • Behandlungsphase (Tag -35 bis Tag 1): ◦ a) Knochenmarkentnahme oder b) Mobilisation des peripheren Blutes mit anschließender Apherese ◦ Isolierung von CD34+-Zellen aus a) einer Knochenmarkprobe oder b) mobilisiertem peripherem Blut für Behandlungs- und Ersatzprobe ◦ Konditionierung mit Busulfan (SMAC) ◦ Verabreichung der OTL-200-Infusion • Follow-up-Phase: Nachbeobachtung nach der OTL-200-Infusion für 8 Jahre¹⁾ Die Studie ist abgeschlossen.
Population	Wesentliche Einschlusskriterien • Dokumentierte biochemische und molekulare MLD-Diagnose, basierend auf ARSA-Aktivität unterhalb des Normbereichs und Identifikation zweier krankheitsverursachender ARSA-Allele. • Erfüllen eines der folgenden Kriterien: a) Älteres Geschwisterkind mit MLD, dessen Alter beim Auftreten der Symptome ≤ 6 Jahre war. Einstufung als LI, EJ oder intermediär (LI/EJ) auf Basis des Alters beim Auftreten der Symptome beim älteren Geschwisterkind und dessen ARSA-Genotyp: ◦ LI: Auftreten von Symptomen beim älteren Geschwisterkind im Alter ≤ 30 Monate; Genotyp typischerweise 0/0. ◦ EJ: Auftreten von Symptomen beim älteren Geschwisterkind im Alter > 30 Monate und ≤ 6 Jahre; Genotyp typischerweise 0/R. ◦ Intermediär: Auftreten von Symptomen beim älteren Geschwisterkind im Alter ≤ 6 Jahre, aber nicht möglich, das ältere Geschwisterkind eindeutig als LI oder EJ zu charakterisieren. - Oder b) Falls die MLD bei einem präsymptomatischen Kind ohne älteres betroffenes Geschwisterkind diagnostiziert wurde, die dem ärztlichen Personal zur Verfügung stehenden Daten aber auf das Vorliegen einer früheinsetzenden MLD-Form hinweisen und das Kind ≤ 6 Jahre alt war, konnte das Kind nach

Charakteristikum	Beschreibung
	Diskussion und mit Zustimmung des klinischen Monitors als geeignet eingestuft werden. - Abwesenheit der folgenden krankheitsbezogenen Symptome: a) Verzögerungen in der Entwicklung der Geh- und Stehfähigkeit und Auffälligkeiten bei der neurologischen Untersuchung oder b) dokumentierte neurologische Symptome der MLD in Verbindung mit kognitiver, motorischer oder verhaltensbezogener Funktionsbeeinträchtigung oder Regression. Das Auftreten von Symptomen wurde vor Behandlungsbeginn erneut überprüft. Wesentliche Ausschlusskriterien - Infektion mit HIV, Tuberkulose, Hepatitis B/C. - Maligne neoplastische Erkrankung oder dokumentierte Vorgeschichte eines hereditären Krebsyndroms. - Myelodysplasie, typische für das myelodysplastische Syndrom oder die akute myeloische Leukämie zytogenetische Veränderungen oder andere schwere, hämatologische Erkrankungen. - Beteiligung an anderen Studien. - Vorherige allogene hämatopoetische Stammzelltransplantation und Nachweis von residualen allo genen Zellen. - Vorherige Genterapie. - Symptomatischer Herpes Zoster, nicht ansprechend auf gezielte Behandlung. - Endorgandysfunktionen oder andere schwere Krankheiten, die eine Studienteilnahme nach Einschätzung der klinischen Prüfperson nicht gestattet.
Intervention und Zahl der Patientinnen und Patienten	Gescreente Testpersonen: keine Angabe Eingeschlossene Testpersonen: N = 5 Personen, die OTL-200 erhielten: N = 5
Ort und Zeitraum der Durchführung; Datenschnitte	Studienzentren 1 Studienzentrum in Mailand, Italien Studienzeitraum - Erste Testperson erste Visite: 16.01.2017 - Letzte Testperson letzte Visite: 03.03.2025 Datenschnitte Interim-Datenschnitt 1 05.12.2018²⁾ (GBA-Erstbewertung)³⁾ Interim-Datenschnitt 2 22.04.2022⁴⁾ Finaler Datenschnitt 26.05.2025⁵⁾ Es liegt ein gemeinsamer Studienbericht (19.12.2025) mit der Studie HE 205029 (siehe Tabelle 7) vor.
Endpunkte gemäß Studienprotokoll	- Neurologische Manifestationen der Erkrankung (IQ, BSID, WPPSI oder WISC-III in Abhängigkeit des Alters, Nine-Hole Peg Test, Schulanwesenheit). - Grobmotorische Funktionen (GMFM und GMFC-MLD). - Beteiligung des peripheren Nervensystems, beurteilt mittels ENG-Untersuchung, anhand der Nervenleitgeschwindigkeit einzelner Nerven und des NCV-Index. - Beteiligung des zentralen Nervensystems, beurteilt mittels Gehirn-MRT. - EEG zur Beurteilung der funktionellen Reifung der Gehirnaktivität. - Biologische Wirksamkeitsendpunkte

Charakteristikum	Beschreibung
	• ARSA-Aktivität • Engraftment der transduzierten Zellen • Lentivirus-Integration • Unerwünschte Ereignisse.

¹⁾ Mit Studienstart der Langzeit-Nachbeobachtungsstudie OTL-200-10 im Dezember 2022 hatten die Teilnehmenden zudem die Möglichkeit, in diese überzugehen.

²⁾ Anlass war ein Antrag auf Marktzulassung.

³⁾ Datenschnitt wurde in der vorangegangenen Nutzenbewertung zu OTL-200 berücksichtigt [10].

⁴⁾ Anlass war ein Antrag auf Zulassung eines biologischen Arzneimittels (Biologics License Application, BLA).

⁵⁾ Datenschnitt gemäß Modul 4. Dem gemeinsamen CSR der Studie CUP 206258 mit der Studie HE 205029 lässt sich ein konkreter Datenschnitt nicht entnehmen; es ist lediglich ein Datum für die Datenbanksperre angegeben (25.07.2025).

Abkürzungen: ARSA: Arylsulfatase A; BSID: Bayley Scales of Infant and Toddler Development; CSR: Clinical Study Report; EEG: Elektroenzephalographie; EJ: Early Juvenile; ENG: Elektroneurographie; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; LI: Late Infantile; MLD: Metachromatische Leukodystrophie; MRT: Magnetresonanztomographie; NCV: Nerve Conduction Velocity; SMAC: Sub-myeloablative Konditionierung; WISC: Wechsler Intelligence Scale for Children; WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

Protokolländerungen

Das Originalprotokoll ist auf den 06. Januar 2017 datiert. Protokoll-Amendment 1 wurde am 12. Juli 2017 und somit rund 6 Monate nach Studieneinschluss der ersten Testperson (16. Januar 2017) erstellt, gefolgt von Protokoll-Amendment 2 am 31. Juli 2018. Relevante Protokolländerungen sind in Tabelle 6 dargestellt.

Tabelle 6: Für die Nutzenbewertung relevante Protokolländerungen der Studie CUP 206258

Amendment	Wesentliche Änderungen
Version 1 vom 12.07.2017, (Anzahl der bis dahin eingeschlossenen Personen: 3)	• Änderung der Einschlusskriterien, die eine Teilnahme von Patientinnen und Patienten mit präsymptomatischer EJ-MLD und intermediärer MLD erlaubt. Zuvor war lediglich die LI-MLD zulässig, definiert als das Erfüllen von 2 der 3 folgenden Kriterien: ◦ Alter bei Auftreten der Symptome bei älteren Geschwisterkindern < 30 Monate. ◦ 2 krankheitsverursachende Allele bei der Testperson (und, falls zutreffend, beim älteren Geschwisterkind). ◦ Periphere Neuropathie laut ENG. • Mit dem Protokoll-Amendment wurde das Kriterium einer peripheren Neuropathie als Einschlusskriterium entfernt.¹⁾ • Festlegung einer Mindestanzahl an CD34+-Zellen für die Knochenmark-entnahme.²⁾
Version 2 vom 31.07.2018, (Anzahl der bis dahin eingeschlossenen Personen: 5)	Keine relevanten Änderungen.

¹⁾ Es ist unklar, ob die beiden Kinder, die nach Amendment 1 eingeschlossen wurden, eine periphere Neuropathie aufwiesen.

²⁾ Es ist unklar, ob vor dieser Änderung für alle bis dahin eingeschlossenen Testpersonen ausreichend CD34+-Zellen zur Verfügung standen.

Abkürzungen: EJ: Early Juvenile; ENG: Elektroneurographie; LI: Late Infantile; MLD: Metachromatische Leukodystrophie

2.2.5 Studie HE 205029

Tabelle 7: Charakterisierung der Studie HE 205029

Charakteristikum	Beschreibung						
Design und Studienablauf	Bei der Studie HE 205029 handelt es sich um eine Fallserie von im Rahmen einer Hospital Exemption (HE) mit OTL-200 behandelten Personen mit LI-MLD. Die Studie unterteilt sich, in Anlehnung an die Studie 201222, in die folgenden Phasen: • Screeningphase (unklarer Zeitraum): Überprüfung der Patienteneignung • Baselinephase (Ende des Screenings bis Tag -5): Erhebung von Baseline-Messwerten und Knochenmarkentnahme zur Gewinnung von Ersatzstammzellen für Ersatzprobe (ca. 3 Wochen vor Behandlung) • Behandlungsphase (Tag -4 bis Tag 0): ◦ Knochenmarkentnahme ◦ Isolierung von CD34+-Zellen aus einer Knochenmarkprobe ◦ Konditionierung mit Busulfan (SMAC) ◦ Verabreichung der OTL-200-Infusion • Follow-up-Phase: Nachbeobachtung nach der OTL-200-Infusion für 8 Jahre¹⁾ Die Studie ist abgeschlossen.						
Population	Wesentliche Einschlusskriterien • Dokumentierte biochemische und molekulare MLD-Diagnose, basierend auf ARSA-Aktivität unterhalb des Normbereichs und Identifikation zweier krankheitsverursachender ARSA-Allele. • LI-Form, definiert durch das Vorliegen von 2 der folgenden Kriterien: ◦ Alter bei Auftreten der Symptome beim älteren Geschwisterkind < 30 Monate. ◦ 2 krankheitsverursachende (Null [0]) Allele. ◦ Periphere Neuropathie laut ENG. • Präsymptomatisch, definiert als keine neurologische Beeinträchtigung im Rahmen der neurologischen Untersuchung, mit oder ohne Anzeichen der Erkrankung, nachgewiesen durch ENG und MRT. • Das Auftreten von Symptomen wurde vor Behandlungsbeginn erneut überprüft. Wesentliche Ausschlusskriterien • Infektion mit HIV, Hepatitis B/C, Treponema pallidum. • Neoplastische Erkrankung. • Typische für das myelodysplastische Syndrom oder die akute myeloische Leukämie zytogenetische Veränderungen. • Endorgandysfunktionen oder andere schwere Krankheiten, die eine Studienteilnahme nach Einschätzung der klinischen Prüfperson nicht gestattet. • Beteiligung an anderen Studien. • Vorherige allogene hämatopoetische Stammzelltransplantation.						
Intervention und Zahl der Patientinnen und Patienten	<table> <tr> <td>Gescreente Testpersonen:</td><td>keine Angabe</td></tr> <tr> <td>Eingeschlossene Testpersonen:</td><td>N = 3</td></tr> <tr> <td>Personen, die OTL-200 erhielten:</td><td>N = 3</td></tr> </table>	Gescreente Testpersonen:	keine Angabe	Eingeschlossene Testpersonen:	N = 3	Personen, die OTL-200 erhielten:	N = 3
Gescreente Testpersonen:	keine Angabe						
Eingeschlossene Testpersonen:	N = 3						
Personen, die OTL-200 erhielten:	N = 3						
Ort und Zeitraum der Durchführung; Datenschnitte	Studienzentren 1 Studienzentrum in Mailand, Italien Studienzeitraum <table> <tr> <td>Einschluss der ersten Testperson:</td><td>29.12.2015</td></tr> <tr> <td>Letzte Testperson letzte Visite:</td><td>03.03.2025</td></tr> </table>	Einschluss der ersten Testperson:	29.12.2015	Letzte Testperson letzte Visite:	03.03.2025		
Einschluss der ersten Testperson:	29.12.2015						
Letzte Testperson letzte Visite:	03.03.2025						

Charakteristikum	Beschreibung
	Datenschnitte Interim-Datenschnitt 1 05.12.2018 ²⁾ (GBA-Erstbewertung) ³⁾ Interim-Datenschnitt 2 22.04.2022 ⁴⁾ Finaler Datenschnitt 26.05.2025 ⁵⁾ Es liegt ein gemeinsamer Studienbericht (19.12.2025) mit der Studie CUP 206258 (siehe Tabelle 5) vor.
Endpunkte gemäß Studienprotokoll⁶⁾	• Neurologische Manifestationen der Erkrankung (BSID, WPPSI oder WISC-III in Abhängigkeit des Alters). • Grobmotorische Funktionen (GMFM und GMFC-MLD). • Beteiligung des peripheren Nervensystems, beurteilt mittels ENG-Untersuchung (NCV-Index) • Beteiligung des zentralen Nervensystems, beurteilt mittels Gehirn-MRT. • EEG zur Beurteilung der funktionellen Reifung der Gehirnaktivität und des Ausbleibens epileptischer Aktivität sowie VEP & BAER zur Beurteilung der Seh- und Hörbahnen • Biologische Wirksamkeitsendpunkte (u. a. ARSA-Aktivität, Engraftment der transduzierten Zellen) • Unerwünschte Ereignisse

¹⁾ Mit Studienstart der Langzeit-Nachbeobachtungsstudie OTL-200-10 im Dezember 2022 hatten die Teilnehmenden zudem die Möglichkeit, in diese überzugehen.

²⁾ Anlass war ein Antrag auf Marktzulassung.

³⁾ Datenschnitt wurde in der vorangegangenen Nutzenbewertung zu OTL-200 berücksichtigt [10].

⁴⁾ Anlass war ein Antrag auf Zulassung eines biologischen Arzneimittels (Biologics License Application, BLA).

⁵⁾ Datenschnitt gemäß Modul 4. Dem gemeinsamen CSR der Studie CUP 206258 mit der Studie HE 205029 lässt sich ein konkreter Datenschnitt nicht entnehmen; es ist lediglich ein Datum für die Datenbanksperre angegeben (25.07.2025).

⁶⁾ Für jede eingeschlossene Testperson (N = 3) wurde ein separates Studienprotokoll vorgelegt.

Abkürzungen: ARSA: Arylsulfatase A; BAER: Hirnstamm-auditorisch-evozierte Potentiale; BLA: Biologics License Application; BSID: Bayley Scale of Infant Development; CSR: Clinical Study Report; ENG: Elektroneurographie; GMFC-MLD: Gross Motor Function Classification for MLD; GMFM: Gross Motor Function Measure; LI: Late Infantile; MLD: Metachromatische Leukodystrophie; MRT: Magnetresonanztomographie; NCV: Nerve Conduction Velocity; SMAC: Sub-myeloablative Konditionierung; VEP: Visuell evozierte Potentiale; WIPPSI: Wechsler Preschool and Primary Scale of Intelligence; WISC: Wechsler Intelligence scale for children.

Protokolländerungen

Für jede Testperson liegt ein eigenes Studienprotokoll vor, jeweils auf den 19.11.2015 datiert. Für eine der Testpersonen wurde am 31.03.2016 ein Protokoll-Amendment erstellt. Aufgrund fehlender Übersichten können Unterschiede sowohl zwischen den Studienprotokollen der Testpersonen als auch zwischen dem initialen Studienprotokoll und dem Protokoll-Amendment der einen Testperson nicht vollumfänglich beurteilt werden. Als relevante Änderung wurde nur der Wechsel des Busulfan-Konditionierungsschemas von myeloablativer Konditionierung (MAC) (initiales Studienprotokoll) auf sub-myeloablative Konditionierung (SMAC) (Protokoll-Amendment) notiert.

2.2.6 Charakterisierung der Intervention im IDS

Tabelle 8: Charakterisierung der Intervention im IDS (Studien 201222, CUP 207394, CUP 206258 und HE 205029)

Intervention
Maßnahmen zur Herstellung von OTL-200
Knochenmarkentnahme oder Mobilisation des peripheren Blutes (MPB) und Apherese zur Herstellung einer autologen Stammzellreserve (für den Fall eines Transplantatversagens) (Tag -35, Studie HE 205029 ca. 3 Wochen vor Behandlung) • Isolierung von CD34+-Zellen aus einer Knochenmarkprobe oder aus mobilisiertem peripherem Blut (G-CSF (10–12,5 µg/kg/Tag) als Mobilisierungsmittel¹⁾) zur Herstellung einer Ersatzprobe mit Ersatzstammzellen (mindestens 2×10^6 CD34+-Zellen/kg). • Insofern CD34+-Zellen aus mobilisiertem peripherem Blut isoliert werden, schließen sich an die Mobilisation Aphereseverfahren an. • Die autologe Stammzellreserve wird kryokonserviert und im Falle eines Transplantatversagens verabreicht.
Knochenmarkentnahme oder Mobilisation des peripheren Blutes und Apherese (Tag -4) • Entnahme von Knochenmark unter sterilen Bedingungen und unter Vollnarkose aus dem Beckenkamm oder Mobilisation des peripheren Blutes mit anschließender Apherese zur Herstellung einer Behandlungsprobe. • Angestrebte Zielanzahl von CD34+-Zellen vor jeglicher Manipulation: ◦ Studie 201222: $\geq 5 \times 10^6$ CD34+-Zellen/kg ◦ Studie CUP 207394: $\geq 5 \times 10^6$ CD34+-Zellen/kg ◦ Studie CUP 206258: 10×10^6 CD34+-Zellen/kg ◦ Studie HE 205029: $\geq 5 \times 10^6$ CD34+-Zellen/kg • Sterile Lagerung bei 4 °C für 24 Stunden.
Aufreinigung der Stammzellen (Tag -3 bis Tag 0) • Reinigung von HSPC aus mononukleären Zellen mittels positiver Selektion mit immunomagnetischen Beads und einem immunomagnetischen Anreicherungsgerät. • Aussaat gereinigter autologer CD34+-HSPC in mit Retronektin beschichteten Beuteln und Inkubation für 24 Stunden bei 37 °C in einem serumfreien Medium (mit Zytokinen ergänzt). • Hinzugabe des ARSA-LV-Vektors in die Zellsuspension und Inkubation für 16 Stunden über Nacht (erste Transduktionsrunde). • Sammlung, Waschung, Zählung und Resuspendierung der Zellen in einem frischen Medium. • Hinzugabe des ARSA-LV-Vektors in die Zellsuspension nach 12 Stunden und Inkubation für 16 Stunden über Nacht (zweite Transduktionsrunde). • Waschung und Vorbereitung der Zellen für die Verabreichung am Ende des Transduktionsprozesses (Verabreichung an Tag 0).
Verabreichung von OTL-200
Busulfan-Konditionierungsschema (Tag -4 bis Tag -1)
<u>SMAC (Studien 201222 (vor Januar 2014), CUP 207394, HE 205029)</u> • 14 Dosen Busulfan (verabreicht als Infusion i. v. alle 6 Stunden). • 1. Dosis auf Grundlage des Körpergewichts. • Anpassung nach der 5. Dosis in Abhängigkeit des Busulfan-Plasmaspiegels bis zu einer Ziel-AUC von 4.800 µg*h/l.
<u>MAC (Studien 201222 (ab Januar 2014), CUP 206258, HE 205029)</u> • 4 Dosen Busulfan (verabreicht als Infusion i. v. alle 20 bis 24 Stunden). • 1. Dosis auf Grundlage der Körperoberfläche der Testperson und je nach Alter.

Intervention
- Anpassung der nachfolgenden Dosen in Abhängigkeit des Busulfan-Plasmaspiegels bis zu einer kumulativen Ziel-AUC von 85.000 µg*h/l OTL-200 (Tag 0 bzw. Tag 1 in Studie CUP 206258) - Infusion i. v. über 10 bis 20 Minuten - Studie 201222: - Minimale Dosis: 2×10^6 CD34+ HSPC/kg - Zieldosis: 5×10^6 bis 10×10^6 CD34+ HSPC/kg - Maximale Dosis: 20×10^6 CD34+ HSPC/kg - Studie CUP 207394 - Minimale Dosis: 2×10^6 CD34+ HSPC/kg - Zieldosis: 5×10^6 bis 10×10^6 CD34+ HSPC/kg - Maximale Dosis: 25×10^6 CD34+ HSPC/kg - Studie CUP 206258: - Empfohlene Dosis: 2×10^6 bis 30×10^6 CD34+ HSPC/kg - Studie HE 205029: - Empfohlene Dosis: 2×10^6 bis 20×10^6 CD34+ HSPC/kg
Begleitmedikation
Supportive Begleittherapie - Hospitalisierung während der Baselinephase zur Implantation eines zentralen Venenkatheters (falls nicht bereits implantiert) für 3–5 Tage (oder in Abhängigkeit des klinischen Zustands). - Hospitalisierung während der Konditionierungstherapie, der OTL-200-Infusion und der kurzfristigen Nachsorge für durchschnittlich 60 Tage. - Beginnend vor der Konditionierung bis zur Erholung von schwerer Neutropenie (ANC < 500/µl) oder bei Entlassung: orale Dekontamination und Mundpflege nach lokalen Standards. - Pentamidin (vernebelt) als Erstlinienprophylaxe von Pneumocystis jirovecii ab Tag 30 (bzw. Tag 5 in den Studien CUP 207394 und 201222) bis zur vollständigen immunologischen Rekonstitution. - Antibakterielle und antimykotische Prophylaxe gemäß lokalem Standard. - Acyclovir als Herpes-Simplex-Prophylaxe bis zur vollständigen immunologischen Rekonstitution. - Intravenöse Substitution zur Aufrechterhaltung des IgG > 5g/l. - Hydratation von 3 l/m² während der Konditionierungsphase (Tag -4 bis 24 Stunden nach der letzten Dosis von Busulfan) - Bei CMV-positivem Befund: präemptive Therapie mit Ganciclovir oder Foscarnet gemäß lokalem Standard. - Bei EBV-positivem Befund nach der Gentherapie und bei wiederholtem Anstieg der EBV-Last: präemptive Therapie mit Rituximab gemäß lokalem Standard. - Verabreichung von Erythrozyten- und Thrombozytenkonzentraten gemäß lokalem Standard (und in Studie CUP 207394 sofern Thrombozyten < 20.000 und Hb < 8,5). - Antiemetische Prophylaxe und Analgetika gemäß lokalem Standard. - Antiepileptika zur Vermeidung von durch die Busulfan-Gabe ausgelösten Krampfanfällen. - GCS-F-Produkte im Falle einer ANC < 500/µl an Tag 45 (bzw. Tag 60 in Studie CUP 207394) nach Untersuchung des Knochenmarks. - Studien CUP 207394, CUP 206258 und HE 205029: parenterale Ernährung bei schwerem Körpergewichtsverlust oder gemäß klinischer Entscheidung. <u>Zusätzlich in Studie CUP 207394</u> - Magenschutz: Antra 10 mg/Tag, Ranidil 5 ml/Tag. - Diuretika: Lasix, Dopamin und Diamox bei Bedarf. - Vitamin K: Koagulopathieprophylaxe und ggf. Ugurol bei schwerer prolongierter Thrombozytopenie und Mukositis.

Intervention
- Bei Bluthochdruck: Amlodipin +/- ACE-Hemmer und Nifedipin im Falle von Spitzen im Blutdruck. - Bei Anzeichen einer metabolischen Azidose: Bikarbonat-Supplementierung i. v.
Nicht erlaubte Begleitmedikation in Studie 201222²⁾ - Enzyersatztherapie mit ARSA-Protein. - Hämatopoetische Stammzelltransplantation von gesunden Personen. - Jegliche andere Gen-/Zelltherapeutika.
Erlaubte Begleitmedikation (alle Studien) - Verabreichung anderer Medikamente in Abhängigkeit vom klinischen Befund gestattet. - Studie 201222: Bestimmte Begleitmedikamente, die zur Prophylaxe vor Transplantationseingriffen und während des Krankenhausaufenthalts nach der Transplantation verwendet wurden (z. B. Antiemetika, Analgetika, Antiepileptika, orale Dekontaminationsmedikamente und Diuretika) oder zur Sedierung/Vollnarkose gelten als Standardbehandlung im Studienzentrum gemäß den lokalen SOP und wurden nur im eCRF erfasst, wenn sie im Zusammenhang mit einer Begleiterkrankung oder einem UE standen. Antibakterielle, antivirale und antimykotische Prophylaxe und Transfusionen wurden routinemäßig im eCRF erfasst.

¹⁾ In Studie 206258 zusätzlich Plerixafor (0,24 mg/kg, subkutan), um Mobilisierung zu verstärken.

²⁾ Für Studien CUP 207394, CUP 206258 und HE 205029 nicht beschrieben.

Abkürzungen: ACE: Angiotensin Converting Enzyme; ANC: Absolute Neutrophilenzahl; ARSA: Arylsulfatase A; AUC: Area Under the Curve; CMV: Zytomegalievirus; CUP: Compassionate Use Program; EBV: Epstein-Barr-Virus; eCRF: electronic Case Report Form; G-CSF: Granulozyten-Kolonie-stimulierender Faktor; Hb: Hämoglobin; HE: Hospital Exemption; HSPC: Hämatopoetische Stamm- und Vorläuferzellen; IDS: Integrierter Datensatz; i.v.: intravenös; LV: Lentiviral; MAC: Myeloablative Konditionierung; SMAC: Sub-myeloablative Konditionierung; SOP: Standard Operating Procedure; UE: Unerwünschtes Ereignis.

2.3 Endpunkte

In diesem Kapitel werden die Eignung der Endpunkte hinsichtlich Operationalisierung, Patientenrelevanz und Validität beurteilt. Dazu wurden das Herstellerdossier, die zugehörigen Studienberichte, Studienprotokolle, SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt. Alle Endpunkte der eingeschlossenen Studien (siehe Tabelle 1) wurden einer Prüfung unterzogen. Endpunkte, die in der nachfolgenden Tabelle 9 nicht gelistet sind, wurden weder vom pU noch im Rahmen der Nutzenbewertung als patientenrelevant bzw. bewertungsrelevant eingestuft. Tabelle 9 stellt das Ergebnis dieser Bewertung zusammenfassend dar.

Tabelle 9: Zusammenfassung der Endpunktbewertung der Studien 201222, CUP 207394, CUP 206258 und HE 205029

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Gesamtüberleben	Mortalität	Ja	Ja
GMFC-MLD	Morbidität	Ja	Nein ¹⁾
Überleben ohne schwere motorische Beeinträchtigung		Ja	Nein
GMFM-Score ²⁾		Ja	Ja
Neuropsychologische Tests ³⁾		Ja	Nein
Entwicklungsquotient ³⁾		Ja	Nein ⁴⁾

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
ARSA-Aktivität ²⁾		Ja	Nein
Unerwünschte Ereignisse	Sicherheit	Ja	Ja

¹⁾ Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

²⁾ Ko-primärer Endpunkt in der Studie 201222.

³⁾ Die Endpunkterhebung wurde mittels verschiedener Instrumente (BSID, WPPSI und WISC) durchgeführt.

⁴⁾ Die Rücklaufquoten liegen bereits zu Baseline unter 70 %. Daher erfolgt keine Beschreibung und Bewertung des Endpunkts in Kapitel 2.3.2 und der Endpunkt wird in der Nutzenbewertung nicht berücksichtigt.

Abkürzungen: ARSA: Arylsulfatase A; BSID: Bayley Scales of Infant and Toddler Development; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; MLD: Metachromatische Leukodystrophie; pU: pharmazeutischer Unternehmer; WISC: Wechsler Intelligence Scale for Children; WPPSI: Wechsler Intelligence Preschool and Primary Scale of Intelligence.

2.3.1 Mortalität

Gesamtüberleben

Der Endpunkt „Gesamtüberleben“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Todesfälle wurden in allen eingeschlossenen Studien im Rahmen der Sicherheit kontinuierlich ab dem Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis zum Ende der Studienteilnahme (also bis zur letzten Studienvisite) erfasst.

Der Endpunkt „Gesamtüberleben“ (sekundärer Endpunkt) war nicht in den Protokollen der IDS-Studien prädefiniert, kann jedoch dem integrierten SAP des IDS (2019) entnommen werden.

Der pU legt folgende Auswertungen vor:

- Überlebensrate
- Alter zum Zeitpunkt des Todes

Bewertung

Die Operationalisierung des Endpunkts ist nachvollziehbar. Zur Auswahl der Auswertungen siehe Kapitel 2.4.

Patientenrelevanz

Der Endpunkt „Gesamtüberleben“ wird als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen.

Validität

Die Erhebung des Endpunkts wird als valide erachtet.

2.3.2 Morbidität

GMFC-MLD

Der Endpunkt „GMFC-MLD“ wird aufgrund unklarer Validität der Analysen (siehe Kapitel 2.4) nicht in der Nutzenbewertung berücksichtigt. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

Operationalisierung

Beschreibung

Beim GMFC-MLD handelt es sich um ein Klassifikationssystem zur Beschreibung grobmotorischer Funktionen von Kindern mit MLD, die mindestens 18 Monate alt sind. Das Klassifikationssystem umfasst 7 verschiedene Stufen (siehe Abbildung 3). In der Publikation zur Entwicklung und Validierung des Instruments wird die Klassifizierung der Kinder durch ärztliches Personal anhand von schriftlichen Beschreibungen der Kinder (z. B. Krankenakten) vorgenommen [16]. In Modul 4 des Dossiers gibt der pU an, dass ebenfalls Elterninterviews verwendet werden konnten.

Klinisch relevante Stadien bei MLD	
Level 0	Altersentsprechendes, freies Gehen ohne Unterstützung
Level 1	Freies Gehen mit Auffälligkeiten, z. B. Gangunsicherheit
Level 2	Gehen nur mit Unterstützung, freies Gehen nicht möglich (weniger als 5 Schritte)
Level 3	Freies Sitzen und Fortbewegung, z. B. Krabbeln. Gehen auch mit Unterstützung nicht möglich
Level 4 a	Freies Sitzen, aber keine Fortbewegung oder
Level 4 b	Kein freies Sitzen, aber Fortbewegung, z. B. Krabbeln
Level 5	Keine Fortbewegung und kein Sitzen ohne Unterstützung; Kopfkontrolle erhalten
Level 6	Verlust jeder Fortbewegung, keine Kopf- und Rumpfkontrolle mehr erhalten

Abbildung 3: Stufen des Klassifikationssystems zur Beschreibung der grobmotorischen Funktionen bei MLD [27]

Während der Langzeitnachbeobachtung (> 3 Jahre) waren Erhebungen im örtlichen Krankenhaus möglich, sofern eine Testperson das Studienzentrum nicht mehr aufsuchen konnte.

Weitere Informationen zur konkreten Durchführung des GMFC-MLD wurden für die vorliegenden Studien nicht identifiziert.

Prädefiniert im SAP des IDS war eine deskriptive Auswertung des Alters bei Eintritt in jede Kategorie und der Verweildauer in der Kategorie.

Der pU legt in Modul 4 für den IDS folgende Post-hoc-Auswertungen vor:

- Anteil der Personen mit GMFC-MLD-Level ≥ 3 zum Zeitpunkt des letzten Follow-ups
- Alter bis zum GMFC-MLD-Level ≥ 3
- Anteil der Personen mit GMFC-MLD-Level ≥ 5 zum Zeitpunkt des letzten Follow-ups
- Alter bis zum GMFC-MLD-Level ≥ 5

Insofern zum Datenschnitt kein Ereignis vorlag, wurden Patientinnen und Patienten zum Zeitpunkt der letzten GMFC-MLD-Erhebung zensiert.

Bewertung

Die Operationalisierung ist eingeschränkt nachvollziehbar. Angaben zur Endpunkterhebung, z. B. ob die Endpunkterhebung durch Schulungen oder Manuale standardisiert wurde und ob die Daten durch Elterninterviews oder eine klinische Prüfperson während der Studienvisite generiert wurden, fehlen. Da die Studien 201222, CUP 207394, CUP 206258 und HE 205029 in demselben Zentrum durchgeführt wurden, wird jedoch von Datenerhebungen in vergleichbarer Weise ausgegangen.

Patientenrelevanz

Die grobmotorische Funktion wird als unmittelbar patientenrelevant bewertet. Es ist offen, inwieweit die Übergänge in die jeweils nächste GMFC-MLD-Stufe seitens der betroffenen Kinder unabhängig vom Ausgangsniveau als gleichrangig bewertet werden.

Validität

Die Validität des Endpunkts „GMFC-MLD“ wurde im vorherigen Nutzenbewertungsverfahren bereits bewertet [10, 16]. Insgesamt liegen nur wenig aussagekräftige Ergebnisse zur Testgüte und keine Ergebnisse zur Änderungssensitivität des GMFC-MLD vor. Da die Stufen des GMFC-MLD jedoch inhaltlich nachvollziehbare und abgrenzbare Krankheitsstadien repräsentieren, wird eine ausreichend valide Erfassung der motorischen Funktion angenommen. Für die durch den pU angeführte besondere Bedeutung der Aufrechterhaltung eines GMFC-MLD-Levels von < 5 oder < 3 konnten keine adäquaten Referenzen identifiziert werden. Der Verlust der Fähigkeiten, sich selbstständig fortbewegen und selbstständig sitzen zu können (Level ≥ 5), kann jedoch als Übergang in ein schwerwiegendes Krankheitsstadium nachvollzogen werden. Die Analysen unter Verwendung einer Schwelle von ≥ 3 lagen für die vorangegangene Bewertung von OTL-200 nicht vor und waren zudem nicht prädefiniert. Ein Grund für die zusätzliche Anwendung dieser Schwelle kann dem vorliegenden Dossier nicht entnommen werden. Die folgenden Ausführungen in der Nutzenbewertung beziehen sich demnach ausschließlich auf Auswertungen zur Schwelle von ≥ 5 .

Die Endpunkte „GMFM“ und „GMFC-MLD“ erfassen beide die grobmotorische Funktion. Somit ist von einer Überlappung zwischen den beiden Endpunkten und damit einer Doppelerfassung der grobmotorischen Funktion auszugehen. Für den Endpunkt „GMFC-MLD“ liegen jedoch mehrere Unklarheiten hinsichtlich der Validität der statistischen Analysen vor (siehe Kapitel 2.4.), die weitestgehend bereits im vorherigen Verfahren zu OTL-200 diskutiert wurden und darin resultieren, dass die Ergebnisse des Endpunkts auch im vorliegenden Verfahren nicht in der Nutzenbewertung herangezogen werden. Diese Unklarheiten liegen für den Endpunkt „GMFM“ nicht vor, welcher zur Nutzenbewertung herangezogen wird.

Überleben ohne schwere motorische Beeinträchtigung

Der Endpunkt „Überleben ohne schwere motorische Beeinträchtigung“ wird aufgrund der unklaren Validität der Analysen der Teilkomponente „GMFC-MLD“ (siehe Kapitel 2.4) nicht in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Das Überleben ohne schwere motorische Beeinträchtigung ist operationalisiert als das Vermeiden von

- a) dem Verlust der Fähigkeit, sich selbstständig fortbewegen und ohne Unterstützung sitzen zu können, definiert durch ein GMFC-MLD-Level ≥ 5 und
- b) Tod aus jeglicher Ursache.

Angaben zur Operationalisierung der Einzelkomponente „Tod aus jeglicher Ursache“ finden sich in Abschnitt 2.3.1 und der Einzelkomponente „Verlust der Fähigkeit, sich selbstständig fortbewegen und ohne Unterstützung sitzen zu können“ unter dem Endpunkt „GMFC-MLD“.

Der kombinierte Endpunkt ist im finalen SAP (Amendment 2) der Studie 201222 (vom 17.09.2019) erstmals erwähnt.

Der pU legt in Modul 4 folgende Auswertungen vor:

- Anteil überlebende Personen ohne schwere motorische Beeinträchtigung
- Alter zum Zeitpunkt einer schweren motorischen Beeinträchtigung oder Tod

Im Rahmen der Ereigniszeitanalyse „Alter zum Zeitpunkt einer schweren motorischen Beeinträchtigung oder Tod“ waren als Ereignis definiert:

- der früheste Verlust der Fortbewegung und des Sitzens ohne Unterstützung, die Kontrolle des Kopfes ist möglich (GMFC-MLD Level ≥ 5) ODER
- Tod aus jeglichen Gründen.

Patientinnen und Patienten, die zum Zeitpunkt des jeweiligen Datenschnitts nicht verstorben sind oder mit einem GMFC-MLD-Level < 5 eingestuft wurden, wurden zum Datum der letzten adäquaten GMFC-MLD-Erhebung zensiert.

Bewertung

Es handelt sich um einen kombinierten Endpunkt, der sich aus Endpunkten der beiden Kategorien Mortalität und Morbidität zusammensetzt.

Während die Operationalisierung der Einzelkomponente „Tod jeglicher Ursache“ nachvollziehbar ist, verbleiben hingegen Unklarheiten bei der Operationalisierung des Endpunkts „GMFC-MLD“ (siehe GMFC-MLD).

Patientenrelevanz

Die Endpunktkomponente „Mortalität“ wird über den Endpunkt „Gesamtüberleben“ berücksichtigt und als patientenrelevant angesehen. Auch die Aufrechterhaltung grobmotorischer Funktionen wird als unmittelbar patientenrelevant eingeschätzt, jedoch sind die endpunktspezifischen Einschränkungen zum GMFC-MLD zu berücksichtigen (siehe Kapitel 2.3.2 und 2.4).

Validität

Die Erhebung der Teilkomponente „Tod aus jeglicher Ursache“ wird als valide erachtet. Obschon für den GMFC-MLD Einschränkungen bezüglich der Validierung vorliegen (siehe Endpunkt GMFC-MLD), wird das Instrument für die Bewertung grobmotorischer Funktionen in Stufen als hinreichend valide angesehen. Für die durch den pU angeführte besondere Bedeutung der Aufrechterhaltung eines GMFC-MLD-Levels von < 5 oder < 3 konnten keine adäquaten Referenzen identifiziert werden. Der Verlust der Fähigkeiten, sich selbstständig fortbewegen und selbstständig sitzen zu können (Level ≥ 5), kann jedoch als Übergang in ein schwerwiegendes Krankheitsstadium nachvollzogen werden.

Für die Teilkomponente „GMFC-MLD-Level ≥ 5 “ sind die bei dem Endpunkt „GMFC-MLD“ genannten Limitationen hinsichtlich Operationalisierung und Validität der statistischen Analysen zu berücksichtigen, weshalb auch der kombinierte Endpunkt nicht für die Nutzenbewertung herangezogen wird.

GMFM

Der Endpunkt „GMFM“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Aufbau des Instruments

Der GMFM-88 umfasst insgesamt 88 Testaufgaben, die sich an einem gesunden und motorisch normal entwickelten 5-jährigen Kind orientieren und von diesem gewöhnlich problemlos ausgeführt werden können. Die Aufgaben sind, entsprechend den kindlichen motorischen Entwicklungsschritten, in 5 Dimensionen unterteilt:

- Aufgabe A: Liegen und Drehen (17 Aufgaben)
- Aufgabe B: Sitzen (20 Aufgaben)
- Aufgabe C: Krabbeln und Knien (14 Aufgaben)
- Aufgabe D: Stehen (13 Aufgaben)

- Aufgabe E: Gehen, Laufen und Springen (24 Aufgaben)

Die Bewertung jeder einzelnen Testaufgabe erfolgt auf Grundlage von Beobachtungen mit einer 4-Punkte-Ordinalskala (0–3 Punkte):

- 0 = initiiert nicht
- 1 = initiiert (Durchführung von < 10 %)
- 2 = führt Aufgabe teilweise aus (Durchführung von 10 bis 100 %)
- 3 = führt gestellte Aufgabe aus (Durchführung 100 %)

Für den Großteil der Aufgaben sind genaue Bewertungsrichtlinien formuliert, die für die Bewertung herangezogen werden müssen.

Durchführung und Auswertung des Tests

Die vollständige Durchführung der GMFM dauert ungefähr 45 bis 60 Minuten. Das Testhandbuch definiert allgemeine Richtlinien zur standardisierten Durchführung des Tests. Für jedes Item sind maximal 3 Versuche zulässig, wovon der beste Versuch aufgezeichnet wird. Verbale Ermunterungen, der Einsatz von Spielsachen zum Anreiz, die Demonstration der Aufgaben und Hilfestellung beim Einnehmen der Ausgangsstellung, nicht aber bei der Durchführung, sind erlaubt. Es wird empfohlen, die Testung zunächst ohne Hilfsmittel und Orthesen durchzuführen. Im Vorfeld können individuelle Zieldimensionen (A–E) ausgewählt werden, bei denen die größten Veränderungen erwartet werden. Laut Testhandbuch [42] sollten die Testungen ausschließlich von in der Anwendung der GMFM trainierten Personen durchgeführt werden.

Für jede Dimension wird ein Prozentwert errechnet:

$$\text{Dimension (\%)} = \frac{\text{erreichte Punktzahl des Kindes}}{\text{Maximalpunktzahl}} * 100$$

Der GMFM-Gesamtwert ergibt sich aus der Summe der Prozentwerte der Einzeldimensionen A–E dividiert durch die Gesamtzahl der Dimensionen:

$$\text{Gesamtwert (\%)} = \frac{\% A + \% B + \% C + \% D + \% E}{5}$$

Der Wertebereich der GMFM reicht von 0–100 %, wobei ein höherer Wert stärker ausgeprägten grobmotorischen Funktionen entspricht.

Es wurden für den Endpunkt „GMFM“ wie präspezifiziert ausschließlich deskriptive Ergebnisse für das einarmige IDS vorgelegt.

Bewertung

Die Operationalisierung ist eingeschränkt nachvollziehbar. Es fehlen Angaben, ob die Tests gemäß der im Testmanual definierten Richtlinien standardisiert durchgeführt wurden. Zudem bleibt unklar, inwieweit die Testdurchführung in den Studien 201222, CUP 207394, CUP 206258 und HE 205029 aufgrund der Dauer von 45 bis 60 Minuten unterbrochen und zu einem späteren Zeitpunkt fortgeführt werden durfte. Dies ist laut Testmanual [42] zulässig. Angesichts der Durchführung der Studien im selben Zentrum wird jedoch von Testungen in vergleichbarer Weise ausgegangen.

Patientenrelevanz

Die vom GMFM erfassten Aspekte der motorischen Entwicklung sind patientenrelevant.

Validität

Die Validität des Endpunkts „GMFM“ wurde im vorherigen Nutzenbewertungsverfahren bereits bewertet [10]. Insgesamt zeigten sich für den GMFM über alle untersuchten Gütekriterien (interne Konsistenz, Konstruktvalidität, Test-Retest-Reliabilität) gute bis sehr gute Werte, insbesondere für Kinder mit Zerebralparese. Es ist jedoch darauf hinzuweisen,

dass Studien zur Validierung der GMFM in der Indikation MLD nicht vorgelegt wurden. Eine Diskussion des pU zur Übertragbarkeit der GMFM-Testergebnisse von Studien mit Kindern mit Zerebralparese auf die Indikation MLD wurde nicht identifiziert. Dennoch erscheint dies grundsätzlich plausibel. Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass das Instrument Funktionen abfragt, die von einem gesunden, motorisch normal entwickelten 5-jährigen Kind ausgeübt werden können.

Die Endpunkte „GMFM“ und „GMFC-MLD“ erfassen beide die grobmotorische Funktion. Somit ist von einer Überlappung zwischen den beiden Endpunkten und damit einer Mehrfacherfassung der grobmotorischen Funktion auszugehen. Während die Ergebnisse des Endpunkts „GMFC-MLD“ vorliegend nicht berücksichtigt werden (siehe Endpunkt „GMFC-MLD“ oben), wird der Endpunkt „GMFM“ für die Nutzenbewertung herangezogen.

Neuropsychologische Tests

Der Endpunkt „Neuropsychologische Tests“ wird in der Nutzenbewertung in Anbetracht einer nicht nachvollziehbaren Operationalisierung und fehlender Informationen zur Validität nicht dargestellt.

Operationalisierung

Beschreibung

Als Instrumente zur Einschätzung des kognitiven Leistungsvermögens wurden abhängig vom Alter der Patientinnen und Patienten drei neuropsychologische Testverfahren eingesetzt:

- Bayley Scales of Infant and Toddler Development (BSID), Version 2 und Version 3 (Altersgruppe: Säuglinge und Kleinkinder von 1 bis 42 Monaten)
- Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Version 3 (zwei Altersversionen:
 - a: Kinder von 2 Jahren und 6 Monaten bis 3 Jahren und 11 Monaten
 - b: Kinder von 4 Jahren bis 7 Jahren und 3 Monaten
- Wechsler Intelligence Scale for Children (WISC), Version 3 und Version 4 (Altersgruppe: Kinder und Jugendliche von 6 Jahren bis 16 Jahren und 11 Monaten)

Der pU gibt an, dass bei Patientinnen und Patienten mit im Vorfeld bereits bekannten Einbußen im kognitiven Leistungsniveau neuropsychologische Testverfahren zum Einsatz kommen konnten, die für jüngere Altersgruppen konzipiert wurden. Zugleich lässt sich die Aussage finden, dass lediglich jene Fragebögen in die Auswertungen einbezogen wurden, die bei Kindern in einem angemessenen Alter zum Einsatz kamen. In Bezug auf die Tatsache, dass sich die Altersgrenzen überschneiden, in denen die 3 neuropsychologischen Fähigkeitstests zum Einsatz kommen, erläuterte der pU zudem im Stellungnahmeverfahren zur vorangegangenen Nutzenbewertung von OTL-200 [12], dass bei überlappenden Altersbereichen jenes Instrument eingesetzt wurde, das nach Einschätzung des durchführenden Neuropsychologen am besten geeignet war. Entscheidend war, mit welchem Instrument die Patientin bzw. der Patient die Aufgaben voraussichtlich am wahrscheinlichsten bewältigen konnte, um fehlende Daten zu vermeiden. Lagen für eine Patientin oder einen Patienten mehrere ausgefüllte Instrumente vor, wurde das altersgerechte Instrument anhand des auf den nächsten Monat gerundeten Alters der Patientin / des Patienten ermittelt. Waren am selben Tag mehrere Instrumente altersgerecht, wurde für die Berichterstattung das Instrument für die niedrigste Altersgruppe herangezogen. Laut pU war dies nur der Fall, wenn das zunächst vorgesehene Instrument nicht vollständig bearbeitet werden konnte und anschließend ein Instrument für eine jüngere Altersgruppe eingesetzt wurde. Alle neuropsychologischen Tests wurden vom gleichen Neuropsychologen durchgeführt.

Mit dem Nutzendossier legt der pU Ergebnisse zu 5 Indizes vor, die sich aus Untertests der 3 neuropsychologischen Testverfahren zusammensetzen:

Tabelle 10: Übersicht zu neuropsychologischen Untertests zur Indexbildung

Index	Eingeschlossene Untertests
Handlungs-IQ	• BSID-II/-III: Kognitive Skala • WPPSI-III (beide Altersversionen): Handlungsteil • WISC-III/-IV: Wahrnehmungsgebundenes logisches Denken
Verbal-IQ	• BSID-II/-III: Sprach-Skala • WPPSI-III (beide Altersversionen): Verbalteil • WISC-III/-IV: Sprachverständnis
Gesamt-IQ	• WPPSI-III (beide Altersversionen): Gesamt-IQ • WISC-III/-IV: Gesamt-IQ
Verarbeitungsgeschwindigkeits-Index	• WPPSI-III (Altersversion b): Verarbeitungsgeschwindigkeit • WISC-III/-IV: Verarbeitungsgeschwindigkeit
Arbeitsgedächtnis-Index	• WISC-IV: Arbeitsgedächtnis

Abkürzungen: BSID: Bayley Scales of Infant and Toddler Development; WISC: Wechsler Intelligence Scale for Children; WPPSI: Wechsler Preschool and Primary Scale for Intelligence.

Für die Auswertungen der Indizes wurde der Summenwert der jeweiligen Untertests der neuropsychologischen Testverfahren in einen standardisierten Wert mit einem Mittelwert von 100 und einer SD von 15 skaliert. Der pU legt dabei einen IQ von ≤ 55 als Indikator für eine schwere geistige Behinderung und als Responsekriterium fest.

In Modul 4 legt der pU für die 5 in Tabelle 10 genannten Indizes folgende Auswertungen vor:

- Deskriptive Darstellung der Ergebnisse über den Studienverlauf
- Anzahl der Responder (Responseschwelle ≤ 55): Bei Patientinnen und Patienten ohne Baseline-Wert wurde keine Imputation vorgenommen.

Bewertung

Die Operationalisierung des Endpunkts ist nicht nachvollziehbar. Der pU führt keine Begründung für die Auswahl der jeweiligen Untertests für die Gesamt-IQ-Skala und die 4 Indizes an. Da sich die Altersgrenzen überschneiden, in denen die 3 neuropsychologischen Fähigkeitstests zum Einsatz kommen, ist undurchsichtig, welche Untertests ab welchem Alter eingesetzt wurden. Der pU gibt zwar an, dass bei überlappenden Altersbereichen das nach Einschätzung des Neuropsychologen am besten geeignete Instrument eingesetzt wurde. Entscheidend war, mit welchem Instrument die Patientin bzw. der Patient die Aufgaben voraussichtlich am ehesten bewältigen konnte. Dies lässt jedoch subjektiven Spielraum, welcher Test angewendet wird. Da gemäß pU der Bewertungsalgorithmus das verwendete Instrument berücksichtigen würde, würde die Wahl des Instruments die resultierende Punktzahl nicht beeinflussen. So sollte nach Auffassung des pU bspw. ein dreijähriges Kind bei gleicher kognitiver Leistung mit dem BSID (1 Monat bis 42 Monate) und dem WPPSI (Altersversion: 2 Jahre und 6 Monate bis 3 Jahre) vergleichbare Werte erzielen [12]. Da keine vollständigen Manuale vorgelegt wurden und der verwendete Bewertungsalgorithmus der einzelnen Instrumente bzw. deren Untertests nicht bekannt ist, kann dies nicht nachvollzogen werden. Da alle Tests vom gleichen Neuropsychologen im gleichen Studienzentrum durchgeführt wurden, könnten daraus resultierende Unterschiede in Entscheidungen zur Wahl des Instrumentes zwischen Patientinnen und Patienten zumindest reduziert worden sein.

Bei Patientinnen und Patienten mit bekannten kognitiven Einbußen konnte ein für jüngere Altersgruppen konzipiertes Instrument verwendet werden, zugleich sollten lediglich jene Fragebögen in die Auswertungen einbezogen werden, die bei Kindern in einem angemessenen

Alter zum Einsatz kamen. Infolgedessen ist unklar, ob alle mit den Instrumenten erhobenen Daten in die Auswertung eingegangen sind.

Der pU berichtet keine konkreten Angaben, welcher Test wann und bei wie vielen Patientinnen und Patienten eingesetzt wurde. Es liegen zudem ausschließlich Ergebnisse für die durch den pU zusammengeführten Indizes vor, nicht aber für die zugrundeliegenden Instrumente. Darüber hinaus ist unklar, ob die neuropsychologischen Testungen durch die Patientinnen und Patienten ab einem gewissen Alter selbst durchgeführt wurden oder ob, je nach Schweregrad der Erkrankung, generell eine Fremdeinschätzung durch Erziehungsberechtigte oder eine klinische Prüfperson stattfand. Angaben dazu, inwiefern die Testverfahren unter möglichst standardisierten Bedingungen durchgeführt wurden, fehlen. Es ist nicht bekannt, welche Referenzpopulation für die Ermittlung der standardisierten Werte herangezogen wurde. In den Unterlagen des pU ließen sich keine vollständigen Manuale der kognitiven Tests identifizieren.

Der pU wählt einen IQ-Wert von ≤ 55 als Responsekriterium für eine schwere geistige Behinderung. Dem Klassifikationssystem DSM-IV folgend wird ein IQ-Wert von ≤ 55 als eine mittelgradige Intelligenzminderung definiert [1]. Gemäß der aktuellen Version des Klassifikationssystems DSM (DSM-5-TR) werden intellektuelle Beeinträchtigungen (leicht, moderat, schwer und schwerst) inzwischen anhand der individuellen Anpassungsfähigkeit graduiert und nicht anhand von festen IQ-Wert-Grenzen, da die Anpassungsfähigkeit – nicht der IQ – den Unterstützungsbedarf bestimmt [2].

Patientenrelevanz

Kognitive und sprachliche Funktionen sind bei MLD in der Kategorie „Morbidity“ patientenrelevante Endpunkte.

Validität

Aufgrund fehlender Informationen zum Einsatz der neuropsychologischen Tests ist eine Bewertung der Objektivität nicht möglich.

Es liegen keine Angaben zur Validität der vom pU zusammengeführten Indizes vor und es ist unklar, inwiefern die einzelnen Subskalen der verschiedenen Messinstrumente hinreichend vergleichbar sind, um eine Zusammenfassung zu rechtfertigen.

Zusätzlich lagen die Rücklaufquoten für den Gesamt-IQ, den Verarbeitungsgeschwindigkeits-Index und den Arbeitsgedächtnis-Index weitestgehend bereits zu Baseline unter 70 % (außer für den Gesamt-IQ in der frühsymptomatischen Teilpopulation).

ARSA-Aktivität

Der Endpunkt „ARSA-Aktivität“ wird in der Nutzenbewertung aufgrund fehlender Patientenrelevanz und Validität als Surrogatparameter nicht berücksichtigt.

Operationalisierung

Beschreibung

Die Rekonstruktion der enzymatischen Aktivität von ARSA wurde a) in mononukleären Zellen des peripheren Blutes (PBMC), b) in aus Knochenmark stammenden mononukleären Zellen (BM) und c) in der Zerebrospinalflüssigkeit (CSF) gemessen.

Zur Bestimmung der ARSA-Enzymaktivität wurden p-Nitrocatecholsulfat (ARSA_{PBMC} und ARSA_{BM}) und, im Fall einer überschüssigen Probenmenge, 4-Methylumbelliferyl-Sulfat (ARSA_{PBMC}, ARSA_{BM} und ARSA_{CSF}) verwendet.

Bewertung

Die Operationalisierung des Endpunkts ist nachvollziehbar.

Patientenrelevanz

Die ARSA-Aktivität ist ein Laborparameter und wird zur Diagnosestellung genutzt. Laborparameter sind nicht per se patientenrelevant und als Surrogatparameter anzusehen.

Anhand der vom pU eingereichten Unterlagen lässt sich nicht ableiten, ob und warum die ARSA-Aktivität im zugrundeliegenden Anwendungsgebiet einen validen Surrogatendpunkt darstellt und Aussagen zu patientenrelevanten Endpunkten zulässt.

Vor diesem Hintergrund wird der Endpunkt „ARSA-Aktivität“ als nicht patientenrelevant erachtet.

Validität

Aufgrund der fehlenden Patientenrelevanz wird die Validität des Endpunkts nicht beurteilt.

2.3.3 Sicherheit

Unerwünschte Ereignisse

Der Endpunkt „Unerwünschte Ereignisse“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Ein unerwünschtes Ereignis (UE) ist definiert als jedes unerwünschte medizinische Vorkommnis, das einer Person, der die OTL-200-Infusion verabreicht wird, widerfährt, unabhängig davon, ob ein kausaler Zusammenhang mit diesem Produkt vermutet worden ist.

Alle UE wurden gemäß Medical Dictionary for Regulatory Activities (MedDRA) Version 28.0 kodiert. Den Studienunterlagen kann nicht entnommen werden, mittels welcher Version der Common Terminology Criteria of Adverse Events (CTCAE) der Schweregrad der UE klassifiziert wurde.

Sämtliche Vorerkrankungen oder Begleiterkrankungen mit CTCAE-Grad ≥ 2 wurden als potentielle UE in Betracht gezogen, wenn sie:

- a) mit oder nach Unterzeichnung der Einwilligungserklärung abgeklungen sind,
- b) nach Unterzeichnung der Einwilligungserklärung weiterhin fortbestehen oder
- c) mit oder nach Unterzeichnung der Einwilligungserklärung aufgetreten sind.

Symptome der Grunderkrankung mit CTCAE-Grad ≥ 3 wurden ferner als UE berichtet.

Die nachfolgend aufgelisteten UE, die in den ersten 100 Tagen der Behandlungsphase aufgetreten sind, wurden nur als UE dokumentiert, wenn sie, im Falle einer klinischen Manifestation, mit einem CTCAE-Grad ≥ 2 oder, im Falle von Laborparametern, mit CTCAE-Grad ≥ 3 klassifiziert wurden:

- Übelkeit
- Erbrechen
- Stomatitis
- Mukositis
- Diarrhö
- Haarausfall
- Fieber unbekannter Genese bei Neutropenie
- Erschöpfung
- Appetitverlust
- Zytopenie
- Blutung (klinisch)
- Kopfschmerzen

- Hypotonie
- Hypertonie
- Vasomotorische Beschwerden (einschließlich Hitzewallung)
- Schwindel
- Allergische Reaktionen (auf Transfusionen und andere Medikamente)
- Infektionen des Zentralvenenkatheters
- Elektrolytstörungen
- Leber- und Nierenfunktionsstörungen

Die Auswertungen erfolgten, gemäß A-priori-Festlegung innerhalb des SAP des IDS, für die folgenden Zeiträume:

- Vor der Behandlung: Tag ab Unterzeichnung der Einwilligungserklärung bis Tag vor Beginn der Konditionierung mit Busulfan; umfasst Screening (Tag < -35) und Baseline
- Behandlungsphase: Ab Tag des Beginns der Konditionierung bis zum Ende der letzten Gentherapie-Infusion
- Akute Phase: Vom Ende der letzten Gentherapie-Infusion bis 48 Stunden nach Ende der letzten Gentherapie-Infusion
- 3 Monate nach der Gentherapie: Ab 48 Stunden nach Ende der letzten Gentherapie-Infusion bis einschließlich Tag 100
- Kurzzeitnachbeobachtung: Tag 101 bis einschließlich Tag 1.097
- Langzeitnachbeobachtung: > Tag 1.097
- Gesamtnachbeobachtung: Nach der letzten Gentherapie-Infusion

Darüber hinaus bietet der pU in Modul 4 den Analysezeitraum „Beginn der Konditionierung vor der Anwendung von OTL-200 bis zum Zeitpunkt des Datenschnittes“ an.

Ein schwerwiegendes unerwünschtes Ereignis (SUE) ist jedes nach der Verabreichung der Studienmedikation bei jedweder Dosis aufgetretene unerwünschte medizinische Vorkommnis, das:

- tödlich ist,
- lebensbedrohend ist,
- eine stationäre Behandlung oder eine Verlängerung der stationären Behandlung erforderlich macht,
- zu bleibender oder schwerwiegender Behinderung oder Invalidität führt oder
- mit kongenitalen Anomalien oder Geburtsfehlern einhergeht.

Zudem wurde ein medizinisches Ereignis, das die Patientin oder den Patienten gefährden oder ein Eingreifen erforderlich machen könnte, um eines der oben gelisteten Ereignisse zu verhindern, als SUE eingestuft.

Schäden des hepatobiliären Systems, die die Hy's-Law-Kriterien erfüllten (a. ALT oder AST $\geq 3 \times$ ULN und Bilirubin $> 2 \times$ ULN mit ALP $< 2 \times$ ULN oder b. ALT oder AST $\geq 3 \times$ ULN und INR $> 1,5$) und spontane Fehlgeburten wurden stets als SUE beurteilt.

Aus Modul 4 geht zusätzlich hervor, dass UE mit CTCAE-Grad 4 oder 5 grundsätzlich als SUE klassifiziert wurden.

UE, die zum Abbruch der Studienmedikation führten:

- Da OTL-200 gemäß Fachinformation als Infusionsdispersion nur 1-mal verabreicht werden darf, ist gemäß pU ein Abbruch der Studienmedikation nach der Infusion nicht möglich.

Das Auftreten eines Todesfalls wird durch den pU laut Modul 4 mit einem Therapieabbruch gleichgesetzt.

Hospitalisierungen aufgrund von UE

- Hospitalisierungen aufgrund von UE wurden laut Modul 4 erhoben. Allerdings wurde der Endpunkt, den Angaben des pU entsprechend, a priori nicht weiter spezifiziert.
- Basierend auf Daten aus Modul 4 geht hervor, dass die Anzahl an UE, die zu einer Hospitalisierung geführt haben, und die Zeit bis zum ersten UE, das eine Hospitalisierung zur Folge hatte, betrachtet wurden.

Während für die Studien CUP 206258 und HE 205029 eine Nachbeobachtungsdauer von bis zu 8 Jahren vorgesehen war, konnten Teilnehmende der Studien 201222 und CUP 207394 auch darüber hinaus weiterbeobachtet werden. Bei beiden Studien war nach Jahr 8 eine jährliche Sicherheitserhebung vorgesehen.

Bewertung

Die Operationalisierung der UE ist nur in Teilen nachvollziehbar.

Bei der Erfassung von Symptomen der Grunderkrankung mit einem CTCAE-Grad ≥ 3 als UE ist unklar, welche UE als Symptome der Grunderkrankung gewertet wurden und wer diese Einschätzung vornahm. Jene UE wurden seitens des pU weder auf SOC-/PT-Ebene präspezifiziert noch in Modul 4 aufgeführt. Etwaige Missklassifikationen von tatsächlichen UE als Symptome können daher nicht ausgeschlossen werden, wodurch UE mit CTCAE-Grad < 3 untererfasst sein könnten.

Des Weiteren definiert der pU eine Liste an UE (u. a. Übelkeit, Erbrechen, Stomatitis), die in den ersten 100 Tagen der Behandlungsphase nur als UE dokumentiert wurden, wenn sie, im Falle einer klinischen Manifestation, mit einem CTCAE-Grad ≥ 2 oder, im Falle von Laborparametern, mit CTCAE-Grad ≥ 3 klassifiziert wurden. Gemäß Studienprotokoll der pivotalen Studie 201222 sind diese UE im Zuge einer Knochenmarkentnahme mit anschließender Konditionierung mit Busulfan zu erwarten und wurden daher bei einem CTCAE-Schweregrad ≤ 2 bzw. 3 in diesen ersten 100 Tagen von der Erfassung ausgeschlossen. Dies kann nicht nachvollzogen werden, da die genannten klinischen Maßnahmen für die Anwendung von OTL-200 notwendig sind. Somit liegt in den ersten 100 Tagen der Behandlungsphase eine potentielle Untererfassung von UE unter CTCAE-Grad 2 bzw. 3 (für Laborparameter) vor.

Weiterhin gibt der pU an, dass sämtliche Vorerkrankungen oder Begleiterkrankungen mit CTCAE-Grad ≥ 2 unter bestimmten Bedingungen als potentielle UE in Betracht gezogen wurden. Die Formulierung „als potentielle UE in Betracht gezogen“ lässt jedoch offen, in welchem konkreten Fall die entsprechenden Vor- und Begleiterkrankungen tatsächlich als UE gewertet wurden und in welchen Fällen nicht.

Es ist zu beachten, dass die verwendete CTCAE-Version zur Klassifizierung der UE-Schweregrade nicht bekannt ist. Dem Vorgehen des pU, das Auftreten eines Todesfalls mit einem UE gleichzusetzen, das zu einem Therapieabbruch führte, wird nicht gefolgt.

Die Operationalisierung der Hospitalisierungen aufgrund von UE ist eingeschränkt nachvollziehbar. Es ist unklar, wer das Auftreten der Hospitalisierungen gemeldet und als durch UE bedingt beurteilt hat. Zudem lässt sich aus den Ausführungen in Modul 4 nicht ableiten, unter welchen Bedingungen eine Krankenhauseinweisung als Hospitalisierung gewertet wurde:

- Erfüllen jedwede Krankenhauseinweisungen, unabhängig von der Dauer der Aufenthalte (z. B. weniger als 24 Stunden) das Kriterium?
- Werden darunter ausschließlich Fälle gefasst, die (teil-)stationär behandelt werden mussten oder auch Fälle, die im Krankenhaus ambulant behandelt werden konnten?

Für die Studien 201222 und CUP 207394 ist zudem zu beachten, dass zwar eine Nachbeobachtung über Jahr 8 hinaus möglich war, nach Jahr 8 jedoch lediglich eine jährliche

Sicherheitserhebung vorgesehen war. Somit ist fraglich, inwiefern nach Jahr 8 Sicherheitsereignisse in den beiden Studien vollständig erfasst werden konnten.

Für die Nutzenbewertung werden UE für die beiden nachfolgend aufgelisteten Zeiträume als relevant angesehen, um den gesamten Zeitraum ab Unterzeichnung der Einwilligungserklärung bis zum Datenschnitt abzudecken:

- Vor der Behandlung (ab Unterzeichnung der Einwilligungserklärung bis Tag vor Beginn der Konditionierungstherapie).
- Beginn der Konditionierung bis zum Zeitpunkt des jeweiligen Datenschnitts.

Für das aktuelle Dossier reicht der pU für das IDS getrennt nach prä- und frühsymptomatischen Patientinnen und Patienten jedoch lediglich Ergebnisse für den Zeitraum „Beginn der Konditionierung bis zum Zeitpunkt des jeweiligen Datenschnitts“ ein, welche in der Nutzenbewertung dargestellt werden.

Patientenrelevanz

Der Endpunkt „Unerwünschte Ereignisse“ wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen. Die Patientenrelevanz von Laborparametern ist unklar.

Validität

Einschränkend für die Validität der Sicherheitsergebnisse ist eine fehlende Operationalisierung der UE mit CTCAE-Grad ≥ 3 , die als Symptome der Grunderkrankung galten. Es ist aus diesem Grund auch unklar, welche Symptome, die als UE mit CTCAE-Grad < 3 dokumentiert wurden, nicht in der deskriptiven Analyse berücksichtigt wurden. Im Grundsatz ist auch fraglich, warum schwere Symptome als UE gewertet wurden, da diese unabhängig vom Schweregrad der Endpunktkategorie „Morbidity“ zuzuordnen sind. Ausgewählte UE wurden während der ersten 100 Tage nur selektiv im Falle eines erhöhten Schweregrads registriert. Angesichts dieser selektiven Erfassung bis Tag 100 und einer fehlenden Präspezifizierung von Symptomen der Grunderkrankung können UE mit Schweregrad < 3 untererfasst worden sein. Diese potentielle Untererfassung ist bei der Interpretation der entsprechenden Ergebnisse zu UE mit jeglichem Schweregrad zu berücksichtigen.

Der pU argumentiert in Modul 4, dass ein Abbruch der Studienmedikation nach der Infusion nicht möglich sei, da OTL-200 gemäß Fachinformation als Infusionsdispersion nur 1-mal verabreicht werden darf. Das Auftreten von UE könnte jedoch zur Folge haben, dass die Konditionierung nicht begonnen wird oder die OTL-200-Infusion nach der Konditionierung nicht verabreicht wird. Eine Darstellung ebendieser Therapieabbrüche wegen UE erfolgt seitens des pU nicht.

Hospitalisierungen aufgrund von UE werden in der Nutzenbewertung nicht separat dargestellt, da davon ausgegangen wird, dass UE, die eine stationäre Behandlung erforderlich machten, auch als SUE erfasst wurden und folglich in die Auswertung der Ergebnisse zu SUE eingingen.

2.3.4 Erhebungszeitpunkte

Eine Übersicht der Erhebungszeitpunkte der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte findet sich in Tabelle 11 und Tabelle 12.

Tabelle 11: Erhebungszeitpunkte der berücksichtigten Endpunkte in den Studien 201222 und CUP 207394

Studienvisite	S ¹⁾	B ²⁾	Tag		Monat																
Endpunkt			28	60	3	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96 ³⁾
Mortalität ⁴⁾	kontinuierlich																				
GMFM	-	x	-	-	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Sicherheit	kontinuierlich																				

¹⁾ Die Screening-Phase ist spätestens mit dem Beginn der Behandlungsphase (Entnahme der Stammzellreserve, Tag -35) abgeschlossen.

²⁾ Spätester Messzeitpunkt vor Entnahme des Knochenmarks (Entnahme erfolgte an Tag -4, also entspricht Baseline Tag -5), einschließlich Screening-Erhebungen.

³⁾ Nach Studienjahr 8 sollten die Teilnehmenden in der pivotalen Studie einmal jährlich per Brief, E-Mail, Telefon oder auf anderem verfügbarem Wege kontaktiert und zu ausgewählten Endpunkten befragt werden. Die Erfassung der UE war dabei obligatorisch, die Erfassung des Endpunkts „GMFM“ hingegen optional. In der Studie CUP 207394 konnten die Studienpersonen ebenfalls nach Jahr 8 weiterbeobachtet werden; Endpunkte wurden dabei nicht näher spezifiziert. Den Studienunterlagen kann für den Zeitraum nach Jahr 8 keine konkrete Präspezifikation entnommen werden, welche Endpunkte weiterhin erhoben werden sollten.

⁴⁾ Im Rahmen der Sicherheit erhoben.

Abkürzungen: B: Baseline; GMFM: Gross Motor Function Measure; S: Screening; UE: Unerwünschte Ereignisse.

Tabelle 12: Erhebungszeitpunkte der berücksichtigten Endpunkte in den Studien CUP 206258 und HE 205029

Studienvisite	S ¹⁾	B ²⁾	Tag		Monat																
Endpunkt			28	60	3	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96
Mortalität ³⁾	kontinuierlich																				
GMFM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Sicherheit	kontinuierlich																				

¹⁾ Die Screening-Phase ist spätestens mit dem Beginn der Behandlungsphase (Entnahme der Stammzellreserve, Tag -35) abgeschlossen.

²⁾ Spätester Messzeitpunkt vor Entnahme des Knochenmarks (Entnahme erfolgte an Tag -4, also entspricht Baseline Tag -5), einschließlich Screening-Erhebungen.

³⁾ Im Rahmen der Sicherheit erhoben.

Abkürzungen: B: Baseline; GMFM: Gross Motor Function Measure; S: Screening.

2.4 Statistische Methoden

Mit dem vorliegenden Dossier reicht der pU Studienberichte sowie dazugehörige Protokolle und SAP ein für

- die pivotale Studie 201222 sowie
- die Härtefallprogramme und Hospital-Exemption-Studien CUP 207394, CUP 206258 und HE 205029.

Die pivotale Studie 201222 und die Härtefallprogramme und Hospital-Exemption-Studien (CUP 207394, CUP 206258 und HE 205029) wurden vom pU zu einem Integrated Data Set (IDS) zusammengefasst. Für das IDS wurde ebenfalls ein eigener SAP eingereicht, welcher auf den 17.09.2019 datiert. Es liegen keine Amendments dieses SAP vor. Die Zusammenfassung der genannten Studien zu einem IDS wurde in der vorangegangenen Bewertung als adäquat eingestuft, da alle Studien im selben Behandlungszentrum durchgeführt wurden und sowohl

hinsichtlich der Studiendurchführung als auch der Studienpopulation von einer hinreichenden Vergleichbarkeit ausgegangen wurde. Dieser Einschätzung wird für die vorliegende Nutzenbewertung gefolgt und das IDS wird für die Nutzenbewertung als maßgebliche Datengrundlage angesehen, da es alle relevanten Studien umfasst.

Entsprechend den Befristungsaufgaben reicht der pU die Wirksamkeitsergebnisse getrennt für prä- und fröhsymptomatische Patientinnen und Patienten ein. Gemäß Fachinformation erfolgt eine vergleichbare Unterteilung in Kinder mit LI/EJ-MLD ohne klinische Manifestation bzw. Kinder mit EJ-MLD mit frühzeitiger klinischer Manifestation der Erkrankung (siehe Kapitel 1 Fragestellung). Daher ist diese Unterteilung auch Grundlage für die Nutzenbewertung.

Neben den Studien aus dem IDS reicht der pU Ergebnisse für verschiedene darauf basierende indirekte Vergleiche zu natürlichen Verlaufskontrollen der Studie TIGET-NHx ein. Keiner dieser indirekten Vergleiche wird als ausreichend valide angesehen, weshalb die Ergebnisse nicht für die Nutzenbewertung herangezogen werden. Weitere Informationen zu den indirekten Vergleichen und die Begründung für das Nicht-Heranziehen für die Nutzenbewertung sind Kapitel 2.6 zu entnehmen.

Die folgenden Ausführungen im gesamten Kapitel beziehen sich somit auf die einarmigen Ergebnisse des IDS.

Analysepopulationen

IDS (Studie 201222, CUP 207394, CUP 206258 und HE 205029)

Folgende für die Nutzenbewertung relevante Analysepopulationen wurden im SAP des IDS vom pU definiert:

- **ITT-Population:** Alle Patientinnen und Patienten, die die Ein- und Ausschlusskriterien der Studie erfüllten, für die eine unterzeichnete Einwilligungserklärung vorlag und die OTL-200 erhalten haben.
- **Sicherheitspopulation:** Alle Patientinnen und Patienten, die in eine der 4 Studien eingeschlossen wurden und OTL-200 erhalten haben.

Es muss berücksichtigt werden, dass die vom pU definierte ITT-Population nicht alle Personen berücksichtigt, die in eine der IDS-Studien eingeschlossen wurden und die möglicherweise die Vorbereitungen auf die OTL-200-Behandlung schon begonnen hatten, aber nicht mit OTL-200 behandelt wurden.

Datenschnitte

In der vorherigen Nutzenbewertung zu OTL-200 [10] lagen für die Studien des IDS Datenschnitte zwischen Januar und Dezember des Jahres 2018 vor.

Für die aktuelle Nutzenbewertung liegen folgende, aktualisierte Datenschnitte vor:

- Studie 201222: 30.06.2025
- Studie CUP 207394: 30.06.2025
- Studie CUP 206258: 26.05.2025
- Studie HE 205029: 26.05.2025

Für die Studien des IDS beträgt die zusätzliche Studiennachbeobachtungsdauer somit zwischen 6 und 7 Jahren. Die Rekrutierung der pivotalen Studie war zum Zeitpunkt der letzten Datenschnitte aus 2018 bereits abgeschlossen. Zudem entspricht die aktuelle Stichprobengröße des IDS der Stichprobengröße des IDS der Vorbewertung. Somit scheinen keine neuen Personen in die weiteren Studien des IDS aufgenommen worden zu sein.

Es ist zu beachten, dass die oben genannten, in Modul 4 vom pU angegebenen Datenschnitte nicht anhand der vorgelegten Studienberichte verifiziert werden konnten. Den Studienberichten der Studien des IDS können keine konkreten Datenschnitte entnommen werden.

Präspezifizierte Subgruppenanalysen

Für die in der vorliegenden Nutzenbewertung berücksichtigten Ergebnisse wurden keine relevanten Subgruppenanalysen vorgelegt.

Fehlende Werte und Daten-Imputation

Gemäß SAP des IDS wurden fehlende Daten weder für die deskriptiven Ergebnisse noch für die inferenzstatistische Analyse imputiert.

Es wurden Ersetzungsregeln für fehlende Datumsangaben für Ereigniszeitanalysen formuliert.

Wenn das Datum der Studienvsiste unvollständig war, in der grobmotorische Funktionen mittels der GMFC-MLD klassifiziert wurden, wurde bei einem fehlenden Tag stattdessen der 15. des Monats und bei einem fehlenden Tag und Monat der 30. Juni herangezogen.

Wenn das Datum des Todestages fehlte, wurde folgender Umgang gewählt:

- Für den Fall, dass der Tag zum Zeitpunkt des Todes nicht dokumentiert wurde, wurde der erste Tag des jeweiligen Monats angenommen.
- Für den Fall, dass der Monat zum Zeitpunkt des Todes nicht dokumentiert wurde, wurde der Monat Januar gewählt.
- Für den Fall, dass das Datum, das ersatzweise angenommen wurde, vor dem letzten Tag lag, für den bekannt war, dass die Patientin oder der Patient noch lebte, oder für den Fall, dass das Jahr nicht angegeben wurde, wurde der Tag nach dem letzten Kontakt gewählt.

Einschätzung der einarmigen statistischen Auswertungen

Gesamtüberleben

Die Ereigniszeitanalyse „Alter zum Zeitpunkt des Todes“ ist definiert als Zeitraum von Geburt bis zum Tod aus jeglicher Ursache. Patientinnen und Patienten, die zum Zeitpunkt des Datenschnitts nicht verstorben sind, wurden zum Zeitpunkt des letzten Kontakts zensiert. Die Analyse zum Alter zum Zeitpunkt des Todes kann nicht nachvollzogen werden. Bei der Operationalisierung des Überlebens von Geburt bis zum Tod jeglicher Ursache ist nicht ersichtlich, wie lange die Teilnehmenden jeweils nach Studieneinschluss bzw. Behandlung überlebten. Falls bspw. eine Person im Alter von 6 Jahren eingeschlossen wird und zeitnah verstirbt, wird für diese Person dennoch eine Überlebensdauer von 6 Jahren angegeben, obwohl der Großteil von diesem Zeitraum vor Studien- und Behandlungsbeginn lag. Die einarmige Ereigniszeitanalyse wird nicht für die Nutzenbewertung herangezogen.

GMFC-MLD

Gemäß den Angaben in den Studienunterlagen wird der GMFC-MLD erst ab einem Alter von 18 Monaten angewandt, da das GMFC-MLD-Level auf der Fähigkeit zu Gehen basiert. Den Baseline-Angaben ist zu entnehmen, dass bei 16 von 20 Patientinnen und Patienten der präsymptomatischen Teilpopulation kein GMFC-MLD-Level zu Baseline erhoben wurde. Für diese Personen beginnt die für den Endpunkt auswertbare Beobachtung daher erst mit der ersten vorliegenden GMFC-MLD-Bewertung und kann entsprechend kürzer sein als bei Personen mit einer Bewertung zu Baseline.

Insgesamt fehlen bei den statistischen Analysen zum Endpunkt „GMFC-MLD“ Informationen zu fehlenden Werten und im Zuge der Ereigniszeitanalysen zu Zensierungsgründen. Inwiefern informative Zensierungen vorliegen könnten, ist nicht bekannt.

- Responderanalyse

Zu der Responderanalyse zum Anteil der Patientinnen und Patienten mit „GMFC-MLD-Level ≥ 5 zum Zeitpunkt des letzten Follow-ups“ ist das „letzte Follow-up“ nicht näher spezifiziert. Unklar ist, ob sich dies auf den letzten verfügbaren Wert oder die letzte geplante Studienvisite bezieht. Ersteres kann in patientenindividuell kurzen und patientenübergreifend stark variierenden Beobachtungszeiten resultieren, falls im Verlauf der Studien fehlende Werte auftreten. Bei Bezugnahme auf die letzte geplante Studienvisite ist ebenfalls der Anteil fehlender Werte und zusätzlich der Umgang damit nicht bekannt. Gemäß Modul 4 soll zwar bei fehlenden Werten eine „Worst-Case-Imputation“ durchgeführt werden, bei welcher ein fehlender Wert als Non-Response gewertet wird. Dabei ist zum einen unklar, ob dies für alle Arten von fehlenden Werten (inkl. aufgrund von Todesfällen) galt. Zum anderen kann dies jedoch insbesondere inhaltlich nicht nachvollzogen werden, da die Angaben im Zuge der Auswertungen implizieren, dass eine Response ab einem Level 5 vorlag und höhere Werte beim GMFC-MLD eine größere Einschränkung der grobmotorischen Fähigkeiten bedeuten. Demnach entspräche die Wertung von fehlenden Werten als Response einer „Worst-Case-Imputation“ und nicht – wie in Modul 4 beschrieben – die Wertung von fehlenden Werten als Non-Response.

- Ereigniszeitanalyse

Der pU legt zudem eine Ereigniszeitanalyse zum Alter bis zum GMFC-MLD-Level ≥ 5 vor. Für beide Teilpopulationen wird ein hoher Anteil von Zensierungen berichtet: in der präsymptomatischen Patientengruppe wurden 18 von 20 Personen und in der frühsymptomatischen Gruppe 4 von 9 Personen zum Zeitpunkt ihres letzten verfügbaren GMFC-Scores zensiert. Zensierungsgründe wurden jedoch nicht berichtet.

Andere Analysen, wie beispielsweise 2 prädefinierte deskriptive Auswertungen (Alter bei Eintritt in die jeweilige Kategorie sowie Verweildauer in jeder Kategorie) wurden nicht für die nutzenbewertungsrelevanten Teilpopulationen (prä- und frühsymptomatische Patientinnen und Patienten) vorgelegt.

Analog zum Endpunkt „Gesamtüberleben“ wird die einarmige Ereigniszeitanalyse zum Alter bis zum GMFC-MLD-Level ≥ 5 als nicht adäquat angesehen (siehe oben).

Aufgrund der beschriebenen Limitationen werden die Ergebnisse des Endpunkts nicht für die Nutzenbewertung berücksichtigt. Aus Transparenzgründen erfolgt eine Darstellung der folgenden Responderanalyse im Anhang:

Anteil der Personen mit GMFC-MLD-Level ≥ 5 zum Zeitpunkt des letzten Follow-ups

Überleben ohne schwere motorische Beeinträchtigung

Aufgrund aller Limitationen, die den Teilendpunkt „GMFC-MLD“ (siehe oben) betreffen, werden auch die Ergebnisse des kombinierten Endpunkts „Überleben ohne schwere motorische Beeinträchtigung“ nicht in der vorliegenden Nutzenbewertung dargestellt.

2.5 Verzerrungspotential auf Studien- und Endpunktebene

Da die Studien 201222, CUP 207394, CUP 206258 und HE 205029 ($\triangleq$ IDS) unkontrolliert waren, ist von einem hohen Verzerrungspotential auf Studien- und Endpunktebene auszugehen.

2.6 Indirekte Vergleiche

Ein indirekter Vergleich mit der natürlichen Verlaufskohorte TIGET-NHx unter Verwendung verschiedener Analysepopulationen wurde im Studienprotokoll und im SAP der Studie 201222 prädefiniert und geplant. Zudem liegt ein Studienprotokoll für die Studie TIGET-NHx vor; eigenständige Studienunterlagen zum indirekten Vergleich wurden jedoch nicht vorgelegt.

Der pU reicht jedoch mehrere indirekte Vergleiche ein, in denen mit OTL-200 behandelte Patientinnen und Patienten aus den Studien des IDS und/oder der Cryo-Studie 205756 mit historischen Kontrollen aus der TIGET-NHx-Studie gegenübergestellt werden:

- A Geschwisterkinderanalyse pivotale Studie 201222 vs. TIGET-NHx (MSAS)
- B Frühsymptomatische Personen des IDS vs. TIGET-NHx (ITT-Population)
- C **Geschwisterkinderanalyse IDS vs. TIGET-NHx (MSAS)**
- D Geschwisterkinderanalyse IDS + Cryo-Studie 205756 vs. TIGET-NHx (MSAS)

Die eingereichten indirekten Vergleiche erfolgten somit überwiegend anhand des Matched-Sibling-Analysis-Set (MSAS), das gemäß SAP des IDS alle Patientinnen und Patienten der ITT-Population umfasst, die unbehandelte Geschwister in der NHx-Studie haben, sowie die entsprechenden unbehandelten Geschwister aus der NHx-Studie. Analysen innerhalb des MSAS werden daher im folgenden „Geschwisterkinderanalysen“ genannt.

Für die Ableitung des Zusatznutzens in Modul 4 zieht der pU die Geschwisterkinderanalyse heran, die Kinder im IDS mit ihren Geschwistern in der NHx-Studie vergleicht (siehe oben indirekter Vergleich C). Da für die Nutzenbewertung das IDS als maßgebliche Datengrundlage angesehen wird (siehe Kapitel 2.4), wird der Fokus im Folgenden primär auf diese Geschwisterkinderanalyse des IDS vs. TIGET-NHx gelegt. Der Vergleich bezieht sowohl prä- als auch frühsymptomatische Patientinnen und Patienten ein. Die überwiegende Mehrheit der im MSAS berücksichtigten Patientinnen und Patienten des IDS war jedoch sowohl zum Zeitpunkt des Studieneinschlusses (94 %) als auch zum Zeitpunkt der Gentherapie (88 %) präsymptomatisch, sodass innerhalb dieses Vergleiches für die Teilpopulation der frühsymptomatischen Kinder kaum Daten vorliegen.

Der Vergleich D, der die Studie 205756 (Cryo) zur kryokonservierten Formulierung von OTL-200 einbezieht, wird hingegen aufgrund von Hinweisen auf eine ergebnisgesteuerte Berichterstattung in Bezug auf die Cryo-Studie nicht für die Nutzenbewertung berücksichtigt (vgl. Tabelle 1).

In Anbetracht der dünnen Datenlage für die Teilpopulation b) des Anwendungsgebietes wären zudem verwertbare vergleichende Daten zu dieser Teilpopulation grundsätzlich wünschenswert. Zum indirekten Vergleich B zu frühsymptomatischen Personen des IDS vs. TIGET-NHx (ITT-Population) wird daher am Ende des Kapitels 2.6 gesondert ausgeführt.

2.6.1 Geschwisterkinderanalyse IDS vs. TIGET-NHx (indirekter Vergleich C)

Datenquellen

Auswahl der Vergleichspopulation

Eine systematische Recherche nach Registern mit potentiell geeigneten historischen Vergleichspopulationen wurde nicht durchgeführt. Aufgrund der Seltenheit der MLD ist zwar von einer begrenzten Anzahl geeigneter Datenquellen auszugehen, dennoch bleibt unklar, ob sämtliche relevanten Vergleichspopulationen identifiziert und berücksichtigt wurden. Positiv wird bewertet, dass die TIGET-NHx-Studie im selben Zentrum wie die Studien des IDS durchgeführt wurde. Dies erhöht die Wahrscheinlichkeit für eine Vergleichbarkeit der Studien hinsichtlich der Datenerhebung, der supportiven Behandlung und der allgemeinen Studienabläufe.

Studie TIGET-NHx zum natürlichen Verlauf

Die vom San Raffaele Telethon Institute for Gene Therapy (SR-TIGET) durchgeführte Natural-History-Study (NHx) TIGET-NHx hatte das Ziel, den natürlichen Krankheitsverlauf bei Patientinnen und Patienten mit MLD zu charakterisieren. Sie wurde in Verbindung mit dem klinischen Entwicklungsprogramm von OTL-200 durchgeführt und schloss seit 2004 Patientinnen und Patienten ein. Für die Klassifikation als LI oder EJ wurden in der NHx-Studie dieselben Definitionen wie im klinischen Entwicklungsprogramm verwendet. Die Datenerhebung und die Auswahl der Endpunkte waren gemäß Angaben im SAP des IDS in vergleichbarer Weise zu den Studien des IDS konzipiert. Dabei waren keine fachärztlichen oder instrumentellen Untersuchungen nach einem streng festgelegten Zeitplan erforderlich. Die NHx-Studie wurde im selben Studienzentrum durchgeführt wie die Studien des IDS. Somit wurden auch die relevanten Endpunkte bei den mit OTL-200 behandelten Kindern und bei den Patientinnen und Patienten der NHx-Studie durch dasselbe klinische Team erhoben und ausgewertet. Ergänzend wurden in der NHx-Studie retrospektive Daten herangezogen, um den Krankheitsverlauf der NHx-Patientinnen und -Patienten möglichst umfassend rekonstruieren zu können und dadurch altersangepasste Vergleiche zwischen behandelten und unbehandelten Patientinnen und Patienten zu ermöglichen. Die Teilnehmenden der NHx-Studie erhielten außer BSC keine weitere Therapie.

Für die Daten der NHx-Studie wurde ein Datenschnitt vom 1. November 2021 verwendet. Zum Zeitpunkt dieses Datenschnitts waren die meisten NHx-Probanden gemäß Studienunterlagen verstorben oder befanden sich auf GMFC-MLD-Stufe 6.

Bewertung der verwendeten Vergleichspopulation

Es wurden, laut Modul 4, jene Patientinnen und Patienten in die Studie TIGET-NHx eingeschlossen, die für eine Teilnahme an den Studien 201222, CUP 207394, CUP 206258 und HE 205029 nicht geeignet waren, aber im selben Zentrum betreut wurden. Die Feststellung der MLD-Form in den Studien des IDS und der NHx-Studie erfolgte anhand derselben Kriterien:

- a) Alter bei Symptombeginn des älteren Geschwisterkinds (präsymptomatische LI-Form) oder das Alter bei Symptombeginn der Testperson oder des älteren Geschwisterkinds (prä- oder frühsymptomatische EJ-Form)
- b) Genotyp
- c) Vorliegen einer peripheren Neuropathie (mittels ENG bestätigt)

Da das Vorhandensein einer peripheren Neuropathie als Kriterium zur Diagnosestellung bei beiden MLD-Formen herangezogen wurde, waren für die Einteilung der Patientinnen und Patienten in die LI- oder EJ-MLD-Gruppe das Alter des älteren Geschwisterkinds (präsymptomatische LI- und EJ-Form) oder der Testperson (frühsymptomatische EJ-Form) beim Auftreten erster Symptome sowie der Genotyp ausschlaggebend. Vom pU wird nicht ausgeführt, inwieweit der Genotyp in der Studie TIGET-NHx, in die das erste Kind im Jahr 2004 eingeschlossen wurde, ähnlich valide bestimmt werden konnte, wie in den Studien 201222 (2010), CUP 207394 (2013), CUP 206258 (2017) und HE 205029 (2015) und ob demnach eine potentielle Fehlklassifikation ausgeschlossen werden kann.

Alle Patientinnen und Patienten des Kontrollkollektivs wurden allerdings, unabhängig davon, welche MLD-Form (LI oder EJ) diagnostiziert wurde, zu Studieneinschluss als symptomatisch klassifiziert. Jedoch sind LI-Patientinnen und -Patienten mit klinischer Manifestation der Erkrankung zum Zeitpunkt ihres Einschlusses in die Studie TIGET-NHx nicht vom Anwendungsgebiet umfasst. Bei der gematchten Geschwisterkinderanalyse des IDS (also IDS vs. NHx) trifft dies auf 73 % im Kontrollarm zu.

Die Studienpopulationen der Studie TIGET-NHx und des IDS unterscheiden sich nicht nur hinsichtlich ihres Symptomstatus (prä- und frühsymptomatisch), sondern auch in Bezug auf grobmotorische Funktionen (siehe hierzu auch Anhang Tabelle 25). Der mittlere GMFM-Wert zu einer nicht näher definierten Baseline (weitere Ausführungen hierzu weiter unten, im Folgenden bezeichnet als „Matchingbaseline“) lag im Vergleichsarm (TIGET-NHx) mit 21 % (SD: 30 %) numerisch deutlich niedriger als im Interventionsarm (IDS) mit 61 % (SD: 26 %). Unter den Patientinnen und Patienten der gematchten Geschwisterkinderanalyse konnten 29 % der behandelten Kinder zu Matchingbaseline noch selbstständig gehen (entspricht Level 0 und 1), während für 71 % keine Matchingbaseline-Werte vorlagen. Unter den unbehandelten Kindern der NHx-Studie befanden sich 27 % in Level 1, während für 73 % keine Matchingbaseline-Werte vorlagen. Die Verteilung der GMFC-MLD-Werte scheint somit vergleichbar; aufgrund des hohen Anteils an fehlenden Beurteilungen kann dies jedoch nicht abschließend beurteilt werden.

Im Kontrollarm lag das mediane Alter mit 43 Monaten (min; max: 21; 85) zu Diagnosestellung und 72 Monaten (min; max: 35; 176) zu Studieneinschluss deutlich höher als im IDS (zu Diagnosestellung: 21 Monate (min; max: 2; 132); zu Studieneinschluss: 24 Monate (min; max: 7; 136)). Dieser Altersunterschied ist dadurch erklärbar, dass die Patientinnen und Patienten aus dem Interventionsarm ältere Geschwister hatten, die in der Studie TIGET-NHx betreut wurden und bei denen daher frühzeitig auf die MLD gescreent werden konnte. Zusätzlich geht aus den Studienunterlagen hervor, dass alle Patientinnen und Patienten mit LI-MLD (n = 13) oder EJ-MLD (n = 4) des IDS im präsymptomatischen Stadium diagnostiziert wurden, nachdem ältere Geschwister Symptome entwickelt und infolgedessen eine MLD-Diagnose erhalten hatten. Bei Kindern im historischen Kontrollarm erfolgte die Diagnosestellung dahingegen erst nach Auftreten erster Symptome.

Die Beobachtungsdauer betrug ab dem Zeitpunkt der ersten Datenerhebung im Kontrollarm (TIGET-NHx) im Median 5,4 Jahre (min; max: 2; 18). Somit wurde mindestens eine Person, unter Berücksichtigung der prospektiven und retrospektiven Erhebungen, bis zum Datenschnitt der Studie TIGET-NHx (01. November 2021) über einen Zeitraum von 18 Jahren nachbeobachtet. Im Gegensatz dazu wurden die Patientinnen und Patienten im Interventionsarm (IDS) im Median 10,6 Jahre (min; max: 8; 15) nachbeobachtet. Die Beobachtungszeiten unterscheiden sich demnach erheblich zwischen Interventions- und historischem Kontrollarm.

Im IDS und in der Studie TIGET-NHx liegen ähnlich hohe Anteile an Patientinnen und Patienten mit LI- (IDS: 76 %, TIGET-NHx: 73 %) und EJ-MLD (IDS: 24 %, TIGET-NHx: 27 %) vor. Das Alter bei Symptombeginn ist gemäß pU ein wichtiger prognostischer Faktor für den Erkrankungsverlauf [12], der gemäß pU nach dem Matching auf numerischer Ebene vergleichbar gewesen sei zwischen den Geschwisterkindern des IDS (medianes Alter 18 Monate mit min; max: 15; 75) und der TIGET-NHx (medianes Alter 18 Monate mit min; max: 15; 75). Hierbei ist jedoch zu beachten, dass der Symptombeginn für die präsymptomatischen Patientinnen und Patienten des IDS, die einen Großteil der behandelten Kinder des MSAS ausmachen (94 %), anhand des Alters zu Symptombeginn der älteren Geschwisterkinder geschätzt wurde. Es ist unklar, ob die Schätzung des Alters bei Symptombeginn noch weitere Faktoren berücksichtigte. Sofern bei den präsymptomatischen, im MSAS eingeschlossenen Kindern des IDS das Alter bei Symptombeginn ausschließlich auf Grundlage des Alters bei Symptombeginn der älteren Geschwisterkinder geschätzt wurde, ist die berichtete Vergleichbarkeit zwischen den Studienarmen nur eingeschränkt aussagekräftig. In diesem Fall beruht das für die IDS-Kinder angesetzte Alter bei Symptombeginn unmittelbar auf dem Alter bei Symptombeginn ihrer älteren Geschwisterkinder der Studie TIGET-NHx, sodass keine unabhängige Vergleichsgröße vorliegt.

Weiterhin erscheint die Annahme, dass Patientengruppen, im Falle einer Gleichverteilung der MLD-Formen, eines vergleichbaren Alters zu Symptombeginn und ohne krankheitsmodifizierende Therapie bis zum Tag der Verabreichung von OTL-200, einen ähnlichen Krankheitsverlauf zeigen, zwar im Grundsatz plausibel. Insofern vor Verabreichung

der OTL-200-Infusion allerdings eine Selektion von Patientinnen und Patienten mit prognostisch günstigerem Profil erfolgte, ist diese Annahme hingegen nicht vertretbar. Da, gemäß den vorliegenden Angaben des pU, Kinder, die mit Beginn der Studie 201222 im Jahr 2010 prä- oder frühsymptomatisch waren, in die Studien 201222, CUP 207394, CUP 206258 und HE 205029 eingeschlossen wurden und die Studienmedikation OTL-200 erhalten konnten, wohingegen Kinder mit schwereren kognitiven und motorischen Beeinträchtigungen ab diesem Zeitpunkt weiterhin in die Studie TIGET-NHx aufgenommen wurden, liegt ein eindeutiger Selektionseffekt vor.

Die Daten der TIGET-NHx-Studie umfassen sowohl Quer- als auch Längsschnittdaten. Während für einen Teil der Patientinnen und Patienten Erhebungen zu mehreren Zeitpunkten vorlagen, waren für andere lediglich Daten zu einem einzelnen Zeitpunkt verfügbar. Zugleich wurden für den Endpunkt GMFC-MLD, neben prospektiven Studiendaten, auch retrospektive Daten verwendet, die vor Studieneinschluss erhoben wurden (ermittelt durch Krankenakten oder Elterninterviews). Es bleibt unklar, wie hoch der Anteil der Patientinnen und Patienten ist, für die Längsschnittdaten retrospektiv erfasst wurden, und ob die Daten primär aus Krankenhausakten oder mittels Elternbefragung erhoben wurden. Insgesamt lässt sich nicht beurteilen, ob die Datenquellen, auf dessen Grundlage die retrospektive Datenerhebung erfolgte, die für die Untersuchung relevanten Daten in der erforderlichen Qualität enthielten.

Patientinnen und Patienten der Studie TIGET-NHx erhielten, laut Modul 4, eine rein unterstützende Behandlung (BSC). Welche Therapieoptionen davon umfasst waren, wird vom pU weder für die Studie TIGET-NHx noch für die Studien 201222, CUP 207394, CUP 206258 und HE 205029 ($\triangleq$ IDS) definiert. Da Kinder in der Studie TIGET-NHx ihre MLD-Diagnose im Median mit 43 Monaten (min; max: 21; 85) erhielten, Kinder im IDS dagegen mit 21 Monaten (min; max: 2; 132), konnte die unterstützende Behandlung in der historischen Kontrolle erst in einem höheren Alter erfolgen als die OTL-200-Infusion in Kombination mit symptombezogener, supportiver Therapie im Interventionsarm. Es wird vom pU nicht diskutiert, inwieweit eine unterstützende Behandlung einen relevanten Einfluss auf die Ergebnisse der in den Studien erhobenen Endpunkte haben kann und ob sich ebenjener Einfluss je nach Symptomstatus zu Beginn einer rein unterstützenden Behandlung (früh- oder symptomatisch) unterscheiden kann. Je stärker BSC die Studienergebnisse der erhobenen Endpunkte beeinflusst und je stärker dieser Einfluss mit einer zeitlich früheren Behandlung assoziiert ist, desto eher kann der durch den indirekten Vergleich ohne Brückenkomparator ermittelte Effekt zugunsten von OTL-200 verzerrt sein. Zwar spricht die Durchführung der NHx-Studie im selben Studienzentrum wie die Interventionsstudien dafür, dass die supportive Versorgung grundsätzlich vergleichbar war. Da die Kinder in die historische Kontrolle jedoch bereits ab dem Jahr 2004 eingeschlossen wurden, in die Studien 201222, CUP 207394, CUP 206258 und HE 205029 hingegen erst ab 2010, kann nicht ausgeschlossen werden, dass sich die supportiven Therapieoptionen im Zeitverlauf und somit auch der potentielle Einfluss auf die Studienergebnisse nicht unwesentlich verändert hat.

Identifikation, Auswahl und Erhebung von Confoundern

Literaturrecherche und Auswahl

Informationen zu einer (systematischen) Literaturrecherche zur Identifikation relevanter Confounder können den Studienunterlagen nicht entnommen werden.

Bewertung der verwendeten Confounder

Das Matching erfolgte anhand der MLD-Form (LI und EJ) sowie anhand der Altersspanne der behandelten Patientinnen und Patienten zur Follow-up-Visite in Jahr 2. Der pU benennt die MLD-Form, das Alter bei Symptombeginn sowie den Genotyp als wichtige prognostische

Faktoren [12]. Hinsichtlich der Verteilung der MLD-Formen erscheinen die beiden Behandlungsarme auf deskriptiver Ebene vergleichbar; zur Beurteilung der Vergleichbarkeit des Alters bei Symptombeginn und des Genotyps liegen keine oder keine adäquaten Daten vor. Aufgrund fehlender Angaben zum methodischen Vorgehen bei der Identifikation der Confounder lässt sich zudem nicht beurteilen, ob sämtliche relevanten Confounder berücksichtigt wurden. Die weiterhin bestehenden relevanten Unterschiede im Symptomstatus (prä- vs. frühsymptomatisch) sowie in der kognitiven und motorischen Symptomlast sprechen dafür, dass durch die gewählten Matching-Variablen keine ausreichende Struktureleichheit zwischen den Behandlungsarmen erreicht wurde.

Statistische Methoden für den indirekten Vergleich

Adjustierung für Confounder

Das Matching von Patientinnen und Patienten, die die OTL-200-Infusion erhielten, und Kindern der Studie TIGET-NHx erfolgte gemäß Modul 4 anhand der MLD-Form und des Lebensalters, das nach 2- und/oder 3-jähriger Nachbeobachtung ab Gabe der OTL-200-Infusion zum Zeitpunkt einer beliebigen Studienvisite innerhalb der jeweiligen Altersspanne der behandelten EJ- oder LI-Gruppe liegen musste. Es fand demnach keine paarweise Zuordnung der Patientinnen und Patienten statt. Aus den Unterlagen des pU geht hervor, dass für die Bestimmung der Untergrenze der Altersspanne eine Abweichung von bis zu 3 Monaten zulässig war. Kinder der Studie TIGET-NHx konnten somit bis zu 3 Monate jünger sein als das jüngste Kind aus den Studien 201222, CUP 207394, CUP 206258 und HE 205029 ($\triangle$ IDS). Als Obergrenze wurde das zum jeweiligen Zeitpunkt höchste Alter im IDS gesetzt. Es wird in Modul 4 durch den pU beispielhaft ausgeführt, dass Kinder mit LI-MLD der Studie TIGET-NHx, im Falle einer Altersspanne zu Jahr 2 von 43 bis 55 Monaten in der Teilpopulation aller behandelten LI-Patientinnen und -Patienten im IDS, zum Zeitpunkt einer beliebigen Visite im Alter zwischen 40 und 55 Monaten sein mussten. Lagen für eine unbehandelte Patientin / einen unbehandelten Patienten mehrere in Frage kommende Untersuchungszeitpunkte vor, wurde gemäß Modul 4 der früheste vollständige Datensatz zum Vergleich herangezogen.

Demnach erfolgte kein vollumfängliches statistisches Matching mit Adjustierung für mehrere vorab identifizierte potentielle Confounder, wie beispielsweise im Rahmen eines Propensity-Score-Matchings. Vereinfacht dargestellt wurde für die Kinder der Studie TIGET-NHx eine beliebige Studienvisite herangezogen, zu der ihr Alter innerhalb der auf Basis der behandelten Kinder zu Jahr 2 bzw. Jahr 3 nach Gentherapie festgelegten Altersspanne lag. Eine paarweise Zuordnung einzelner Kinder der NHx-Studie zu einzelnen OTL-200-behandelten Kindern mit identischem Alter erfolgte dabei nicht. Zusätzlich wurde die MLD-Form berücksichtigt. Weitere potentielle Confounder wurden bei der Bildung der Vergleichspopulation nicht berücksichtigt.

Den vorliegenden Unterlagen ist nicht zu entnehmen, wie groß die Ausgangspopulation für das Geschwisterkinder-Matching war, d. h. wie viele Kinder in den beiden Studienarmen grundsätzlich ein Geschwisterkind im jeweils anderen Studienarm hatten und damit für die gematchte Geschwisterkinderanalyse infrage kamen. Daher bleibt unklar, welchen Anteil die letztlich gematchten Geschwisterkinder an der grundsätzlich für das Matching infrage kommenden Population ausmachten und welche Geschwisterkinder aus welchen Gründen nicht gematcht wurden. Darüber hinaus ist nicht nachvollziehbar, wie sichergestellt wurde, dass die gematchte Population tatsächlich verwandte Geschwisterpaare enthielt, sofern keine 1:1-Zuordnung vorgenommen wurde.

In Modul 4 gibt der pU an, dass ein Matching für die Untersuchungszeitpunkte 2 und 3 Jahre nach Behandlung anhand der jeweiligen Altersspanne aller behandelten Patientinnen und Patienten mit LI- bzw. EI-MLD zu diesen Zeitpunkten präspezifiziert war. Aus den vorliegenden

Angaben geht jedoch nicht eindeutig hervor, ob für jeden Untersuchungszeitpunkt ein erneutes Matching vorgesehen war und somit untersuchungszeitpunktspezifische gematchte Vergleichspopulationen gebildet werden sollten. Gleichzeitig wurde/n die Baselineerhebung/en (insbesondere für den NHx-Arm) nicht eindeutig definiert. Besonders für die Charakterisierung der Studienpopulation bleibt daher unklar, auf welchen Zeitpunkt sich die zu Baseline (in der Nutzenbewertung als „Matchingbaseline“ bezeichnet) berichteten Werte der NHx-Gruppe beziehen.

Positivität, Überlappung, Balance

Im Zuge der Geschwisterkinderanalyse IDS vs. TIGET-NHx bestehen wesentliche Unsicherheiten hinsichtlich der zentralen Annahmen der kausalen Inferenz (Positivität, Überlappung und Balance).

So liegen bspw. keinerlei Informationen zum Ausgangsset von allen Geschwisterkindern des IDS und TIGET-NHx vor sowie ob und aus welchen Gründen Geschwisterkinder nicht gematcht werden konnten. Abschließende Einschätzungen zur Positivität und Überlappung der Ausgangspopulationen sind daher nicht möglich. Laut Modul 4 wurden jedoch jene Patientinnen und Patienten in die Studie TIGET-NHx eingeschlossen, die für eine Teilnahme an den Studien des IDS nicht geeignet waren, aber im selben Zentrum betreut wurden. Dies spricht gegen die Positivitätsannahme für die zugrundeliegenden Ausgangs- und gematchten Populationen. Entsprechend zeigt sich auch, dass bereits zu Studieneinschluss 73 % der Kinder im gematchten TIGET-NHx-Arm eine LI-MLD mit Frühsymptomatik aufwiesen und demnach nicht der Zulassungspopulation von OTL-200 entsprachen.

Informationen zum Ausgangsset von allen Geschwisterkindern wären jedoch auch deswegen von Bedeutung, da u. a. anhand der Altersspanne der behandelten Patientinnen und Patienten im IDS zur Follow-up-Visite in Jahr 2 (anstatt anhand der Altersspanne zu Baseline) gematcht werden sollte. Dieses Kriterium wird kritisch gesehen und könnte mit einer ergebnisgesteuerten Berichterstattung einhergehen, da es darauf hindeutet, dass Personen, die vor der Follow-up-Visite in Jahr 2 im OTL-200-Arm bereits verstorben sind ($n = 3$) oder nicht behandelt wurden ($n = 2$), vom Matching im Zuge der Geschwisterkinderanalyse ausgeschlossen wurden, auch wenn sie möglicherweise Geschwister in der TIGET-NHx-Studie hatten und für ein Matching andernfalls in Frage gekommen wären. Im Falle der Anwendung des Kriteriums auf den OTL-200-Arm und den TIGET-NHx-Arm sowie einer Baseline 2 Jahre vor der Follow-up-Visite (im TIGET-NHx-Arm) läge für beide Arme für 2 Jahre ein Immortal-Time-Bias vor.

Das Matchingkriterium zur Altersspanne garantiert kein vergleichbares Alter der gematchten Populationen und es liegen keine Informationen zum Alter der gematchten Populationen zu Matchingbaseline vor. Zum Zeitpunkt der OTL-200-Infusion betrug das mittlere Alter der behandelten Patientinnen und Patienten im gematchten IDS-Arm 34 ± 39 Monate (Median: 18; min; max: 8; 140). Die Altersspanne umfasst demnach im Interventionsarm einen Zeitraum von insgesamt 11 Jahren. Es besteht grundsätzlich die Möglichkeit, dass das Alter der nicht behandelten Kinder zum jeweiligen Untersuchungszeitpunkt zum Großteil nah der Ober- oder Untergrenze der entsprechenden Altersspanne lag.

Da zudem nicht sichergestellt wurde, dass alle relevanten Confounder identifiziert, in beiden Kohorten erhoben und beim Matching berücksichtigt wurden, erscheint vorliegend auch die univariate Vergleichbarkeit beider Kohorten hinsichtlich aller relevanten Patientencharakteristika fraglich. Doch auch aus dem Vergleich der berichteten (univariaten) Patientencharakteristika ergeben sich Unklarheiten, da zur bereits gematchten Population der Geschwisterkinderanalyse des IDS ($N = 17$) und der Studie TIGET-NHx ($N = 15$) nur eingeschränkte Informationen vorliegen. Da sowohl keine klare Baselinedefinition für die

gematchten Geschwisterkinder aus der Studie TIGET-NHx vorliegt als auch zu einem wesentlichen Anteil der Patientencharakteristika (siehe Tabelle 25 im Anhang) nicht eindeutig ersichtlich ist, dass sich die Angaben auf die Matchingbaseline beziehen, kann dieser Aspekt der Balance der beiden Vergleichsgruppen nicht abschließend bewertet werden.

Es zeigt sich jedoch bspw., dass zu Screening/Studieneinschluss bereits 16 (94 %) der gematchten Kinder aus dem IDS präsymptomatisch waren und 1 (6 %) Kind fröhsymptomatisch, während alle gematchten Kinder (n = 15) aus der Studie TIGET-NHx fröhsymptomatisch waren. Zum Zeitpunkt des Studieneinschlusses bestehen damit schon Zweifel an den zentralen Annahmen der kausalen Inferenz, neben der oben genannten Positivität (da die Kinder im TIGET-NHx-Arm gleichzeitig LI-MLD aufwiesen), insbesondere auch an der Balanciertheit. Da Kinder aus der gematchten TIGET-NHx-Gruppe gemäß dem beschriebenen Matchingvorgehen während ihrer gesamten Erhebungszeitspanne theoretisch mehrfach für ein Matching in Frage kommen können, können Studieneinschluss und Matchingbaseline jedoch voneinander abweichen. Infolgedessen scheint nicht gesichert, dass die Geschwisterkinder der gematchten TIGET-NHx-Gruppe zu ihrer Matchingbaseline noch als fröhsymptomatisch kategorisiert werden können oder ggf. bereits progredient sind, was auch auf univariater Ebene gegen die Positivitätsannahme sprechen würde. Da davon auszugehen ist, dass die erkrankungsbedingten Symptome bei Geschwisterkindern in der Studie TIGET-NHx vor dem Matching zu Studieneinschluss schwerwiegender waren, kann auf der anderen Seite nicht ausgeschlossen werden, dass Patientinnen und Patienten infolge des Fortschreitens der Erkrankung verstarben, bevor sie als potentielle Kontrolle in das MSAS eingeschlossen werden konnten. Durch diesen potentiellen Selektionseffekt besteht die Möglichkeit eines Survivorship-Bias im TIGET-NHx-Arm. Insofern Patientinnen und Patienten mit einem prognostisch günstigeren Krankheitsprofil in die vier Studien des IDS eingeschlossen wurden, ergibt sich gleichzeitig hingegen ein Bias zu Gunsten des OTL-200-Arms. Während die univariaten Verteilungen der MLD-Formen zwischen den Vergleichskohorten zwar vergleichbar scheinen und das Lebensalter zum ersten Kontakt angenähert wurde, deuten neben den beobachteten Symptomstatusunterschieden zu Screening/Studieneinschluss auch Unterschiede bei anderen Charakteristika, wie z. B. beim Geschlecht (IDS: 41 % weiblich; TIGET-NHx: 60 % weiblich) und beim medianem GMFM-Score (in %) zu Matchingbaseline (IDS: 65 %; TIGET-NHx: 9,3 %; Range von 0–100 mit hohen Werten als bessere grobmotorische Funktionen) auf eine fehlende Balanciertheit der gematchten Geschwisterkinder hin.

Umgang mit fehlenden Daten

Der Umgang mit fehlenden Werten in den Studien des IDS ist Kapitel 2.4 zu entnehmen.

Gemäß Studienbericht der Studie 201222 sollte grundsätzlich – soweit möglich – ein konservativer Ansatz beim Umgang mit fehlenden Werten gewählt werden: Für die behandelten Patientinnen und Patienten wurde der ungünstigste Fall und für die NHx-Patientinnen und -Patienten der günstigste Fall angenommen. Konkretere Angaben können den Studienunterlagen nicht entnommen werden. Für den Umgang mit fehlenden Werten im Rahmen der Geschwisterkinderanalyse können zudem keinerlei Informationen identifiziert werden.

2.6.2 Zusammenfassende Bewertung der Methodik des indirekten Vergleichs

In der Gesamtschau liegen umfangreiche Unklarheiten zu der aktuell vorliegenden Geschwisterkinderanalyse IDS vs. TIGET-NHx vor. Insbesondere fehlen eine systematische Confounderidentifikation und vollständige Confounderdarstellung, Informationen zum Ausgangsset von allen Geschwisterkindern beider Kohorten sowie eine eindeutige Baselinedefinition für die Geschwisterkinder der Studie TIGET-NHx und die dazugehörigen

Baselinecharakteristika. Es ist ferner unklar, inwiefern sichergestellt war, dass alle gematchten Kinder tatsächliche Geschwister im jeweils gematchten Arm aufwiesen, da ein entsprechendes Kriterium nicht berichtet wurde.

Abgesehen von der fraglichen Gegebenheit der Positivität, wird anerkannt, dass ein umfassendes Matching unter Berücksichtigung zahlreicher Confounder angesichts der ohnehin kleinen Ausgangsstichproben zu einer weiteren erheblichen Reduktion der Stichprobengröße geführt hätte. Dennoch wären eine systematische Identifikation und vollständige Darstellung potentiell relevanter Confounder sowie eine Prüfung ihrer Verteilung zwischen den gematchten Gruppen nach dem Matching zum adäquaten Baseline-Erhebungszeitpunkt zur Beurteilung der Angemessenheit des Matchings sinnvoll gewesen.

Auf Basis der bereits vorliegenden Informationen liegen trotz umfangreicher Unklarheiten bereits Hinweise darauf vor, dass die zentralen Annahmen der kausalen Inferenz (Positivität, Überlappung und Balance) nicht in ausreichendem Maße erfüllt werden. Dies ist u. a. darin begründet, dass in die Studie TIGET-NHx Patientinnen und Patienten eingeschlossen wurden, die für eine Teilnahme an den Studien des IDS nicht geeignet waren und im gematchten TIGET-NHx-Arm 73 % Kinder mit frühsymptomatischer LI-MLD (zu Studieneinschluss, nicht Baseline) verbleiben, die nicht der zugelassenen Indikation von OTL-200 entsprechen und nicht in die Studien des IDS aufgenommen werden sollten. Weitere Hinweise können Kapitel 2.6.1 entnommen werden.

Aufgrund der genannten Limitationen wird der indirekte Vergleich als nicht ausreichend valide betrachtet und für die Nutzenbewertung nicht herangezogen. Aus Transparenzgründen und aufgrund der Größe des Effekts im Rahmen der Ergebnisse der Geschwisterkinderanalyse für den Endpunkt „GMFC-MLD“ erfolgt eine Darstellung des indirekten Vergleichs der gepoolten Population des IDS mit den natürlichen Verlaufskontrollen der TIGET-NHx-Studie im Anhang.

2.6.3 Frühsymptomatische Patientinnen und Patienten des IDS vs. TIGET-NHx (indirekter Vergleich B)

Es ist zu beachten, dass ein Großteil der bereits dargestellten Kritikpunkte (darunter die fehlende Systematik bei der Auswahl der Vergleichskohorte und der Confounder, mögliche Selektionseffekte hinsichtlich des Symptomstatus und der Symptomlast, Unterschiede im Alter bei Diagnosestellung, die unklare Altersverteilung zu den Auswertungszeitpunkten Jahr 2 und 3 sowie Unklarheiten hinsichtlich des möglichen Einflusses potentieller Unterschiede in der BSC) auch auf den indirekten Vergleich B zu frühsymptomatischen Patientinnen und Patienten zutreffen.

Neben den bereits genannten generellen Limitationen des Vergleichs des IDS mit der NHx-Studie erfolgt der Vergleich der frühsymptomatischen Patientinnen und Patienten naiv, d. h. ohne Matching. Dies kann auf die geringe Stichprobengröße der frühsymptomatischen Patientinnen und Patienten im Interventionsarm zurückzuführen sein (N = 8–9, eine Person war zum Screening präsymptomatisch). Die im Anhang dargestellten Charakteristika der beiden Studienarme für diesen Vergleich verdeutlichen jedoch, dass sich die beiden Arme hinsichtlich eines Großteils der berichteten krankheitsbezogenen Faktoren erheblich unterscheiden: darunter das Alter bei Krankheitsbeginn, die MLD-Form, der Symptomstatus sowie die motorische und kognitive Symptomlast (motorisch, kognitiv). Aufgrund der damit fehlenden Strukturgleichheit kann der indirekte Vergleich nicht als valide eingestuft werden; er wird somit nicht für die Nutzenbewertung herangezogen.

3 Ergebnisse der eingeschlossenen Studien

Es wurden insgesamt 31 Patientinnen und Patienten in die Studien 201222, CUP 207394, CUP 206258 und HE 205029 eingeschlossen. Aufgrund eines Widerrufs einer erteilten Einwilligungserklärung vor Behandlungsbeginn in der Studie 201222 umfasst die Gesamtpopulation für die gemeinsame Analyse der Studien 201222, CUP 207394, CUP 206258 und HE 205029 ($\triangleq$ IDS) 30 Patientinnen und Patienten. Eine Person ist anlässlich klinischer Kriterien zu Baseline aus der Studie 201222 ausgeschieden und erhielt die OTL-200-Infusion nicht. Die ITT- und Sicherheitspopulation bilden 29 Patientinnen und Patienten.

3.1 Studiencharakteristika und Studienmedikation

Tabelle 13: Allgemeine Angaben; IDS

IDS Allgemeine Angaben	OTL-200 Präsymptomatisch	OTL-200 Frühsymptomatisch
Eingeschlossen ¹⁾ , n (%)	k. A.	k. A.
Gesamtpopulation ²⁾ , n (%)	k. A.	k. A.
Erhalt der Studienmedikation, n (%)		
Ja	20 (k. A.)	9 (k. A.)
Nein	k. A. ³⁾	k. A. ³⁾
ITT-Population ⁴⁾ , n (%)	20 (k. A.)	9 (k. A.)
Studie 201222	12 (60) ⁵⁾	8 (88,9) ⁵⁾
Studie CUP 207394	0	1 (11,1) ⁵⁾
Studie CUP 206258	5 (26) ⁵⁾	0
Studie HE 205029	3 (15) ⁵⁾	0
Studienabbruch, n (%)		
Tod	1 (5) ⁵⁾	2 (22,2) ⁵⁾
Lost to Follow-up	k. A. ⁶⁾	k. A. ⁶⁾
Beobachtungsdauer ⁷⁾ (Jahre)		
n	20 (100) ⁵⁾	9 (100) ⁵⁾
Mittelwert (SD)	10,1 (3,30)	9,08 (4,75)
Median (min; max)	10,2 (1,14; 15,2)	10,9 (0,64; 13,9)

¹⁾ Es wurden insgesamt 31 Patientinnen und Patienten eingeschlossen. Wie sich diese auf die beiden Teilpopulationen verteilen, kann den Studienunterlagen nicht entnommen werden.

²⁾ Die Gesamtpopulation umfasst die in die vier Studien 201222, CUP 207394, CUP 206258 und HE 205029 eingeschlossenen Patientinnen und Patienten. Eine Person aus der Studie 201222 nahm die Einwilligungserklärung vor Beginn der Behandlung zurück und ist in keinem Quelldatensatz enthalten. Somit bestand die Gesamtpopulation aus 30 Patientinnen und Patienten. Wie sich diese auf die beiden Teilpopulationen verteilen, kann den Studienunterlagen nicht entnommen werden.

³⁾ Bei einer Person aus der Studie 201222 führte eine rasche Krankheitsprogression zwischen dem Screening und einer erneuten Begutachtung zum Zeitpunkt der Baseline-Erhebung zu einem vorzeitigen Studienabbruch, da die Person aufgrund der Progression nicht mehr für die Behandlung geeignet war. Ob diese Person zum Zeitpunkt des Studieneinschlusses prä- oder frühsymptomatisch war, kann den Studienunterlagen nicht entnommen werden.

⁴⁾ Definition, siehe Kapitel 2.4.

⁵⁾ Prozentangaben aus eigener Berechnung, bezogen auf die ITT-Population (präsymptomatisch N = 20; frühsymptomatisch N = 9).

⁶⁾ Den Studienunterlagen kann entnommen werden, dass eine Person in der Studie 201222 die Studie abgebrochen hat; als Grund wird „Lost to Follow-up“ angegeben. Aus den Angaben bei den Ergebnissen zum Gesamtüberleben geht zudem hervor, dass eine Person der frühsymptomatischen Teilpopulation „Lost to Follow-up“ war. Es wird davon ausgegangen, dass dies dieselbe Person ist.

⁷⁾ Gemäß Angaben im SAP des IDS war die Beobachtungsdauer für mit OTL-200 behandelte Patientinnen und Patienten definiert als die Zeit von der Behandlung mit OTL-200 bis zur letzten abgeschlossenen Studienvsiste. Aus welchem Grund der Zeitraum zwischen Unterzeichnung der Einwilligungserklärung und Gentherapie nicht berücksichtigt wurde, ist nicht ersichtlich.

Abkürzungen: IDS: Integrierter Datensatz; ITT: Intention-to-Treat; k. A.: keine Angabe; SAP: statistischer Analyseplan; SD: Standardabweichung.

Tabelle 14: Charakterisierung der Studienpopulation; IDS, ITT-Population

IDS Charakterisierung der Studienpopulation	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
<i>Alter zu (geschätztem) Krankheitsbeginn (Monate)</i> Mittelwert (SD) Median (min; max)	29,4 (20,5) ¹⁾ 18,5 (15; 83) ¹⁾	56,6 (17,1) 60 (26; 83)
<i>Alter zu Diagnosestellung (Monate)</i> Mittelwert (SD) Median (min; max)	14,1 (17,1) 8,83 (0; 60,9)	68,2 (31,6) 68,1 (20,2; 131,9)
<i>Alter zu Screening/Studieneinschluss (Monate)</i> Mittelwert (SD) Median (min; max)	17,3 (16,5) 11,9 (6,57; 61,7)	70,9 (32,6) 69,8 (21,6; 136,2)
<i>Alter zum Zeitpunkt des 1. Kontaktes²⁾ (Monate)</i> Mittelwert (SD) Median (min; max)	17 (16,7) 12 (3,59; 61,8)	71 (32,6) 69,9 (21,6; 136,4)
<i>Alter zum Zeitpunkt der Gentherapie (Monate)</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	12 (60) ⁴⁾ 25,3 (22) 16,3 (7,56; 66,7)	8 (88,9) ⁴⁾ 71,7 (35,2) 70 (23,3; 139,7)
<i>Alter bei letzter Visite (Jahre)</i> Mittelwert (SD) Median (min; max)	11,7 (3,82) 12,5 (2,08; 16,9)	15,3 (5,54) 17 (6,56; 22,4)
<i>Geschlecht, n (%)</i> Weiblich Männlich	7 (35) 13 (65)	6 (67) 3 (33)
<i>Abstammung (genetisch), n (%)</i> Weiß - kaukasisch/europäisch Weiß - arabisch/nordafrikanische Herkunft Asiatisch - südostasiatische Herkunft	16 (80) 3 (15) 1 (5)	8 (89) 1 (11) 0
<i>Ethnie, n (%)</i> Nicht Hispanisch oder Latino Hispanisch oder Latino	19 (95) 1 (5)	9 (100) 0 (0)
<i>Symptomstatus zu Studieneinschluss⁵⁾, n (%)</i> Präsymptomatisch Frühsymptomatisch	20 (100) 0	1 (11) 8 (89)
<i>Symptomstatus zum Zeitpunkt der Gentherapie⁵⁾, n (%)</i> Präsymptomatisch Frühsymptomatisch	k. A.	k. A.
<i>MLD-Form, n (%)</i> Spätinfantil (LI) Frühjuvenil (EJ) ⁶⁾	15 (75) 5 (25)	1 (11) 8 (89)

IDS Charakterisierung der Studienpopulation	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
<i>GMFM-Score (%)</i> ⁷⁾ zu Baseline Mittelwert (SD) Median (min; max)	58,9 (26,3) 61,6 (13,4; 98,6)	84,6 (9,67) 86,6 (71,1; 98,7)
<i>GMFC-MLD</i> ⁸⁾ , n (%) zu Baseline Level 0 Level 1 Level 2 Level 3 Level 4 Level 5 Level 6 k. A. ⁹⁾	4 (20) 0 0 0 0 0 0 16 (80)	2 (22) 7 (78) 0 0 0 0 0 0
<i>ARSA-Aktivität_{PBMC} (nmol/mg/h)</i> zu Baseline n (%) ³⁾ Mittelwert (SD) Median (min; max)	19 (95) ⁴⁾ 27,4 (6,27) 25,8 (25,8; 53,1)	9 (100) 26 (0,73) 25,8 (25,8; 28)
<i>MRT-Score</i> ¹⁰⁾ zu Baseline Mittelwert (SD) Median (min; max)	0,59 (1,41) 0 (0; 4,25)	7,92 (4,32) 10 (0,50; 12)
<i>NCV-Index</i> ¹¹⁾ zu Baseline n (%) ³⁾ Mittelwert (SD) Median (min; max)	19 (95) ⁴⁾ -3,59 (3,22) -3,38 (-10,3; 1,12)	9 (100) -7,01 (2,61) -7,93 (-9,45; -3,11)
<i>Begleitende Therapie zu Studienbeginn, n (%)</i> Ja Nein	k. A.	k. A.
<i>Vorthherapie zu Studienbeginn, n (%)</i> Ja Nein	k. A.	k. A.

¹⁾ Bei präsymptomatischen Patientinnen und Patienten wurde das vorhergesagte Alter zu Krankheitsbeginn auf der Grundlage des Alters beim Auftreten der ersten Symptome bei den älteren Geschwistern der Patientin / des Patienten berechnet. Bei 5 Kindern in der Studie 201222, 2 Kindern der Studie HE 205029 sowie bei einem Kind der Studie CUP 206258 war das genaue Alter beim Auftreten der ersten Symptome beim unbehandelten Geschwisterkind ungewiss und wurde im eCRF als Spanne erfasst; das höhere Alter wurde als voraussichtliches Alter für den Beginn der Symptome beim behandelten Kind herangezogen.

²⁾ Alter zum Zeitpunkt der Baseline-Erhebung von GMFC-MLD, GMFM, NCV und MRT.

³⁾ Eine Angabe von n (%) erfolgt nur bei Charakteristika, für die keine vollständigen Daten vorlagen.

⁴⁾ Prozentangaben aus Eigenberechnung bezogen auf die ITT-Population (präsymptomatisch N = 20; frühsymptomatisch N = 9).

⁵⁾ Daten zur Verteilung des Symptomstatus zum Zeitpunkt der Gentherapie liegen nicht vor. Modul 4 kann entnommen werden, dass eine LI-Testperson zum Zeitpunkt des Screenings noch präsymptomatisch war und zum Zeitpunkt der Gentherapie als frühsymptomatisch eingestuft wurde. Somit wäre diese Person zum Zeitpunkt der Gentherapie nicht vom Anwendungsgebiet umfasst. Es ist nicht eindeutig ersichtlich, ob diese Person für die Analysen des IDS der prä- oder frühsymptomatischen Teilpopulation zugeordnet wurde. Im Studienbericht der pivotalen Studie wurde die Person für die Darstellung der Analysen als präsymptomatisch eingestuft.

⁶⁾ Für die Analysen wurde eine Person der Studie 201222 mit einer intermediären Form der frühjuvenilen Form (EJ) zugeordnet. Ob diese prä- oder frühsymptomatisch war, ist unklar.

⁷⁾ Der Wertebereich der GMFM reicht von 0–100 %, wobei ein höherer Wert stärker ausgeprägten grobmotorischen Funktionen entspricht.

- ⁸⁾ Klassifikationssystem zur Beschreibung grobmotorischer Funktionen von Kindern mit MLD, die mindestens 18 Monate alt sind. Das Klassifikationssystem umfasst 7 verschiedene Stufen (Stufe 0: altersentsprechendes, freies Gehen ohne Unterstützung; Stufe 6: Verlust jeglicher Fortbewegungsfähigkeit sowie jeglicher Kopf- und Rumpfkontrolle).
- ⁹⁾ Da das GMFC-MLD-Level auf der Fähigkeit zu Gehen basiert, kann es in der Regel erst ab einem Alter von mindestens 18 Monaten bestimmt werden. Somit konnte für Kinder mit einem Alter von < 18 Monaten kein Baseline-Wert erhoben werden. Bei Kindern, die jünger sind als 18 Monate, aber schon laufen können, kann das GMFC-MLD-Klassifikationssystem gemäß pU jedoch angewendet werden. Ob alle fehlenden Werte darauf zurückzuführen sind, dass die entsprechenden Kinder < 18 Monate alt waren bzw. noch nicht laufen konnten, kann den Studienunterlagen nicht entnommen werden.
- ¹⁰⁾ Der MRT-Score wurde anhand einer adaptierten Loes' Skala ermittelt, die Subskalen zu Demyelinisierung (0–16 Punkte), b) globaler (0–4 Punkte) und fokaler (0–10,5 Punkte) Hirnatrophie sowie getriggertem Muster ('tigroid pattern') (0–1 Punkte) umfasst. Der Gesamtscore kann Werte von 0 (unauffällig) bis 31,5 (merklich verändert) annehmen. Ein Wert von > 0 gilt als auffällig.
- ¹¹⁾ Der NCV-Index berechnet sich aus dem Durchschnitt von 4 z-Scores, die auf Testungen der Nervenleitgeschwindigkeit des Nervus suralis, Nervus medianus, Nervus fibularis profundus und Nervus ulnaris beruhen. Für die Berechnung der z-Scores wurden, je nach Alter der Patientinnen und Patienten, 3 altersbezogene Normkollektive herangezogen. Der NCV-Index ist normalverteilt mit einem Mittelwert von 0 und einer SD von 0,5. Je höher das Ausmaß der Demyelinisierung, desto kleiner der Index-Wert.
- Abkürzungen: ARSA: Arylsulfatase A; eCRF: electronic Case Report Form; EJ: Early Juvenile; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; IDS: Integrierter Datensatz; ITT: Intention-to-Treat; k. A.: keine Angabe; LI: Late Infantile; MLD: Metachromatische Leukodystrophie; MRT: Magnetresonanztomographie; NCV: Nerve Conduction Velocity; SD: Standardabweichung.

Exposition mit der Studienmedikation

Angaben zur Exposition mit der Studienmedikation sind in Tabelle 15 dargestellt. Alle Patientinnen und Patienten erhielten vor Verabreichung der OTL-200-Infusion eine systemische Konditionierung mit Busulfan.

Der Inhalt jedes OTL-200-Infusionsbeutels sollte, laut den Unterlagen des pU, innerhalb von rund 10 bis 20 Minuten vollständig infundiert werden.

Tabelle 15: Angaben zur Exposition mit der Studienmedikation; IDS, ITT-Population

IDS Exposition mit der Studienmedikation	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
<i>Busulfan-Konditionierungsschema, n (%)</i>		
SMAC	10 (50)	3 (33)
MAC	10 (50)	6 (67)
<i>AUC-Gesamtwert von Busulfan (µg x h/L)</i>		
Mittelwert (SD)	77542,8 (7388,7)	81351,4 (7154)
Median (min; max)	78949,1 (63420; 87741)	84985,9 (70506; 88310)
OTL-200		
<i>Anzahl an CD34+-Zellen (x 10⁶/kg)</i>		
Mittelwert (SD)	12,7 (5,58)	8,14 (2,88)
Median (min; max)	12,5 (4,20; 25,9)	7,10 (3,80; 13,1)

Abkürzungen: AUC: Area Under the Curve; IDS: Integrierter Datensatz; ITT: Intention-to-Treat; MAC: Myeloablative Konditionierung; SD: Standardabweichung; SMAC: Sub-myeloablative Konditionierung.

Protokollverletzungen

Es liegen keine Daten zu den Protokollverletzungen für das gesamte IDS oder getrennt nach prä- und frühsymptomatischen Patientinnen und Patienten vor.

Dem CSR der Studie 201222 lassen sich für die insgesamt 20 prä- und frühsymptomatischen Kinder u. a. Abweichungen in Bezug auf die Studienprozesse und die Untersuchungen/Erhebungen (100 %; u. a. Einhaltung des Untersuchungszeitplans), gefolgt von Abweichungen von der guten klinischen Praxis (80 %; nicht näher definiert) sowie Abweichungen bei der Sicherheitsberichterstattung durch die Prüferärztinnen und -ärzte (60 %; nicht näher definiert) entnehmen. Bei 2 Teilnehmenden traten zudem wesentliche Abweichungen vom Protokoll im Zusammenhang mit der Verabreichung der OTL-200-Infusion auf (Einsatz einer zweiten OTL-200-Charge aufgrund geringer HSPC-Ausbeute sowie eine nach der Verabreichung festgestellte Partikelverunreinigung durch Spritzenmaterial).

Es ist unklar, wie viele Personen in den Studien CUP 206258 und HE 205029 wesentliche Protokollabweichungen aufwiesen. In der Fallstudie CUP 207394 wies der einzige Teilnehmende wesentliche Protokollverletzungen hinsichtlich versäumter Untersuchungen und Abweichungen von der Terminplanung auf.

Aufgrund des einarmigen Studiendesigns liegt ein hohes Verzerrungspotential vor. Zusätzlich wurden zahlreiche Protokollabweichungen berichtet. Der konkrete Einfluss dieser Protokollabweichungen auf die Ergebnisse kann nicht abschließend beurteilt werden.

Begleitmedikation

Es liegen keine Daten zur Begleitmedikation innerhalb des gesamten IDS oder getrennt nach prä- und frühsymptomatischen Patientinnen und Patienten des IDS vor. Ergebnisse auf Einzelstudienebene sind im Anhang dargestellt (Anhang Kapitel „Begleitmedikation auf Einzelstudienebene“). Auswirkungen der Einnahme von Begleitmedikation, die auf das Muskel-Skelett-System wirkt, auf die Ergebnisse des Endpunkts „GMFM“ sind nicht auszuschließen.

Aufgrund des einarmigen Studiendesigns liegt ein hohes Verzerrungspotential vor. Zusätzliche Einflüsse der Einnahme von Begleitmedikation auf die Ergebnisse können nicht abschließend beurteilt werden.

3.2 Mortalität

Es zeigte sich zu den jeweiligen Datenschnitten in den Studien 201222 (30.06.2025), CUP 207394 (30.06.2025), CUP 206258 (26.05.2025) und HE 205029 (26.05.2025) ($\triangleq$ IDS), dass insgesamt 3 von 29 behandelten Patientinnen und Patienten (10 %) mit einer EJ-MLD verstorben sind. Eine dieser Personen (männlich) entstammt der Studie CUP 206258, verstarb zu Monat 14 nach der OTL-200-Infusion aufgrund eines ischämischen Schlaganfalls und war zum Zeitpunkt der Gentherapie präsymptomatisch. Die beiden anderen Personen (beide weiblich) entstammen der Studie 201222, verstarben zu Monat 8 und Monat 15 nach OTL-200 aufgrund einer Dysphagie, infolge des Fortschreitens ihrer Erkrankung und waren zum Zeitpunkt der Gentherapie frühsymptomatisch. Die Kinder wurden auf Wunsch ihrer Eltern nicht mit einer perkutanen endoskopischen Gastrostomie-Sonde ernährt.

Tabelle 16: Ergebnisse für „Gesamtüberleben“ im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); ITT-Population

IDS Gesamtüberleben	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
Todesfälle, n (%)		
Ja	1 (5)	2 (22)
Nein	19 (95)	7 (78)
Kein Ereignis, Studie abgeschlossen	19 (95)	6 (76)
Lost to Follow-up	0	1 (11)
Beobachtungsdauer ¹⁾ (in Jahren), Median (min; max)	10,2 (1,14; 15,2)	10,9 (0,64; 13,9)

¹⁾ Gemäß SAP ist die Beobachtungsdauer definiert als die Zeit von der Gentherapie bis zum Ende der letzten Studienvisite.

Abkürzungen: IDS: Integrierter Datensatz; ITT: Intention-to-Treat; SAP: statistischer Analyseplan.

3.3 Morbidität

GMFM

Da die Rücklaufquoten nach Jahr 4 für beide Teilpopulationen unter 70 % lagen, werden die Ergebnisse bis zum Jahr 4 berichtet.

Tabelle 17: Ergebnisse für „GMFM“ (in %) im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); ITT-Population

IDS GMFM (in %) ¹⁾	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
Gesamtwert		
Gesamtwert zu Baseline n (%)	20 (100)	9 (100)
Mittelwert (SD)	58,9 (26,3)	84,6 (9,67)
Median (min; max)	61,6 (13,4; 98,6)	86,6 (71,1; 98,7)
Gesamtwert zu Jahr 2 n (%)	18 (94,7) ²⁾	6 (85,7) ²⁾
Mittelwert (SD)	83,4 (12,7)	63,3 (29,6)
Median (min; max)	86,3 (57,5; 100)	65,3 (14,7; 98,9)
Gesamtwert zu Jahr 3 n (%)	18 (94,7) ²⁾	6 (85,7) ²⁾
Mittelwert (SD)	86,8 (14,5)	58,3 (29,7)
Median (min; max)	90,6 (40,6; 100)	60,3 (16; 97,9)
Gesamtwert zu Jahr 4 n (%),	15 (78,9) ²⁾	5 (71,4) ²⁾
Mittelwert (SD)	86,3 (20,3)	43,5 (21,9)
Median (min; max)	95 (22,7; 100)	45,8 (10,9; 64)
Dimension A³⁾		
Dimensionswert zu Baseline n (%)	20 (100)	9 (100)
Mittelwert (SD)	90,8 (14,7)	95,4 (6,04)
Median (min; max)	100 (47,1; 100)	100 (86,3; 100)

IDS GMFM (in %)¹⁾	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
Dimensionswert zu Jahr 2 n (%) Mittelwert (SD) Median (min; max)	18 (94,7) ²⁾ 97,6 (4,28) 100 (88,2; 100)	6 (85,7) ²⁾ 88,9 (18,6) 98 (52,9; 100)
Dimensionswert zu Jahr 3 n (%) Mittelwert (SD) Median (min; max)	18 (94,7) ²⁾ 97,9 (4,87) 100 (82,4; 100)	6 (85,7) ²⁾ 86 (18,4) 93,1 (54,9; 100)
Dimensionswert zu Jahr 4 n (%) Mittelwert (SD) Median (min; max)	15 (78,9) ²⁾ 97,8 (8,61) 100 (66,7; 100)	5 (71,4) ²⁾ 78,4 (25) 86,3 (35,3; 100)
Dimension B⁴⁾		
Dimensionswert zu Baseline n (%) Mittelwert (SD) Median (min; max)	20 (100) 76 (28,2) 87,5 (20; 100)	9 (100) 98 (3,20) 100 (91,7; 100)
Dimensionswert zu Jahr 2 n (%) Mittelwert (SD) Median (min; max)	18 (94,7) ²⁾ 97,1 (3,65) 98,3 (88,3; 100)	6 (85,7) ²⁾ 80,8 (32,2) 92,5 (16,7; 100)
Dimensionswert zu Jahr 3 n (%) Mittelwert (SD) Median (min; max)	18 (94,7) ²⁾ 96,6 (9,54) 100 (61,7; 100)	6 (85,7) ²⁾ 79,4 (31,2) 93,3 (20; 100)
Dimensionswert zu Jahr 4 n (%) Mittelwert (SD) Median (min; max)	15 (78,9) ²⁾ 96 (13,7) 100 (46,7; 100)	5 (71,4) ²⁾ 71 (33,5) 85 (16,7; 98,3)
Dimension C⁵⁾		
Dimensionswert zu Baseline n (%) Mittelwert (SD) Median (min; max)	20 (100) 58,7 (36,7) 73,8 (0; 100)	9 (100) 86,5 (13,7) 90,5 (66,7; 100)
Dimensionswert zu Jahr 2 n (%) Mittelwert (SD) Median (min; max)	18 (94,7) ²⁾ 89 (9,39) 89,3 (69,1; 100)	6 (85,7) ²⁾ 66,3 (38,8) 78,6 (0; 100)
Dimensionswert zu Jahr 3 n (%) Mittelwert (SD) Median (min; max)	18 (94,7) ²⁾ 89,8 (16,6) 95,2 (40,5; 100)	6 (85,7) ²⁾ 57,9 (42,7) 70,2 (2,38; 100)
Dimensionswert zu Jahr 4 n (%) Mittelwert (SD) Median (min; max)	15 (78,9) ²⁾ 90 (25,7) 100 (0; 100)	5 (71,4) ²⁾ 44,8 (37,1) 52,4 (2,38; 90,5)

IDS GMFM (in %) ¹⁾	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
Dimension D⁶⁾		
Dimensionswert zu Baseline		
n (%)	20 (100)	9 (100)
Mittelwert (SD)	41,5 (34,2)	78,1 (10,5)
Median (min; max)	37,2 (0; 100)	74,4 (64,1; 94,9)
Dimensionswert zu Jahr 2		
n (%)	18 (94,7) ²⁾	6 (85,7) ²⁾
Mittelwert (SD)	73,7 (22,2)	48,3 (34,9)
Median (min; max)	78,2 (30,8; 100)	41 (2,56; 100)
Dimensionswert zu Jahr 3		
n (%)	18 (94,7) ²⁾	6 (85,7) ²⁾
Mittelwert (SD)	78,6 (19,6)	39,7 (33,9)
Median (min; max)	80,8 (12,8; 100)	30,8 (2,56; 92,3)
Dimensionswert zu Jahr 4		
n (%)	15 (78,9) ²⁾	5 (71,4) ²⁾
Mittelwert (SD)	77,3 (27,3)	18 (16,8)
Median (min; max)	84,6 (0; 100)	15,4 (0; 41)
Dimension E⁷⁾		
Dimensionswert zu Baseline		
n (%)	20 (100)	9 (100)
Mittelwert (SD)	27,3 (30,9)	65,1 (22,7)
Median (min; max)	16 (0; 93,1)	69,4 (33,3; 98,6)
Dimensionswert zu Jahr 2		
n (%)	18 (94,7) ²⁾	6 (85,7) ²⁾
Mittelwert (SD)	59,6 (28,4)	32,2 (36,9)
Median (min; max)	61,8 (6,94; 100)	15,3 (1,39; 94,4)
Dimensionswert zu Jahr 3		
n (%)	18 (94,7) ²⁾	6 (85,7) ²⁾
Mittelwert (SD)	70,9 (28,1)	28,2 (36,6)
Median (min; max)	80,6 (5,56; 100)	16 (0; 97,2)
Dimensionswert zu Jahr 4		
n (%)	15 (78,9) ²⁾	5 (71,4) ²⁾
Mittelwert (SD)	70,5 (34,1)	5,28 (9,54)
Median (min; max)	90,3 (0; 100)	1,39 (0; 22,2)

¹⁾ Der Wertebereich der GMFM reicht von 0–100 %, wobei ein höherer Wert stärker ausgeprägten grobmotorischen Funktionen entspricht.

²⁾ Eigenberechnung: Die Rückläufe wurden berechnet unter Berücksichtigung der aufgetretenen Todesfälle. Insgesamt waren 3 Personen bereits verstorben, eine davon zu Monat 14 (präsymptomatisch) sowie 2 zu Monat 8 und Monat 15 (beide frühsymptomatisch). Somit beziehen sich die Prozentangaben auf N = 19 präsymptomatische und N = 7 frühsymptomatische Personen.

³⁾ Aufgaben zum Liegen und Drehen.

⁴⁾ Aufgaben zum Sitzen.

⁵⁾ Aufgaben zum Krabbeln und Knien.

⁶⁾ Aufgaben zum Stehen.

⁷⁾ Aufgaben zum Gehen, Laufen und Springen.

Abkürzungen: GMFM: Gross Motor Function Measure; IDS: Integrierter Datensatz; ITT: Intention-to-Treat; SD: Standardabweichung.

3.4 Lebensqualität

Es wurden keine Endpunkte der Kategorie „Lebensqualität“ erhoben.

3.5 Sicherheit

Die deskriptive Auswertung der Sicherheitsendpunkte erfolgte auf Basis der Sicherheitspopulation separat für einzelne Studienphasen. Es wurden für die vollständige Darstellung der Endpunkte der Kategorie „Sicherheit“ ab der Unterzeichnung der Einwilligungserklärung bis zu den Datenschnitten in den Studien 201222 (30.06.2025), CUP 207394 (30.06.2025), CUP 206258 und HE 205029 (26.05.2025) zwei Studienphasen herangezogen: a) Zeitraum ab dem Tag der Unterzeichnung der Einwilligungserklärung bis Tag vor Beginn der Konditionierung und b) Zeitraum ab Konditionierung bis zum jeweiligen Datenschnitt. Die Angaben zu den vorgenannten Phasen beziehen sich auf das Neu- oder Wiederauftreten von UE innerhalb der jeweiligen Zeiträume. Somit werden UE, die vor der Konditionierung mit Busulfan auftraten und mit Beginn der Konditionierung nicht abgeklungen sind, nur in der ersten Studienphase erfasst.

Eine konkrete endpunktspezifische Beobachtungsdauer für die unerwünschten Ereignisse im IDS kann den eingereichten Unterlagen nicht entnommen werden. In Modul 4 wird eine endpunkunabhängige mediane Beobachtungsdauer von 10,2 Jahren (min; max: 1,14; 15,2) in der präsymptomatischen und von 10,9 Jahren (min; max: 0,64; 13,9) in der frühsymptomatischen Gruppe berichtet. Gemäß SAP des IDS wird die Beobachtungsdauer berechnet als der Zeitraum ab Gabe der Gentherapie bis zur letzten abgeschlossenen Visite. Ob dabei nur die Gabe der Infusion selbst oder auch die Konditionierung vor der Infusion berücksichtigt wird, ist nicht ersichtlich. Somit ist ebenfalls nicht ersichtlich, ob die genannte endpunkunabhängige Beobachtungsdauer auf die Sicherheitserfassung übertragbar ist.

Tabelle 18: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); Sicherheitspopulation¹⁾

IDS Zusammenfassung der UE <i>Personen mit mindestens einem ...</i>	Studienbeginn bis Konditionierung (Tag -5) ²⁾		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsymptomatisch N = 20 n (%)	OTL-200 Frühsymptomatisch N = 9 n (%)	OTL-200 Präsymptomatisch N = 20 n (%)	OTL-200 Frühsymptomatisch N = 9 n (%)
UE (ergänzend dargestellt)	k. A.	k. A.	20 (100)	9 (100)
UE CTCAE-Grad ≥ 3 ³⁾	k. A.	k. A.	20 (100)	9 (100)
SUE	k. A.	k. A.	17 (85)	9 (100)
UE, das zum Abbruch der Studienmedikation führte ⁴⁾	k. A.	k. A.	k. A.	k. A.

¹⁾ Alle Patientinnen und Patienten, die in eine der 4 Studien eingeschlossen wurden und OTL-200 erhalten haben (präsymptomatisch N = 20; frühsymptomatisch N = 9).

²⁾ Zeitraum ab Tag der Unterzeichnung der Einwilligungserklärung bis zum Tag vor Beginn der Konditionierung mit Busulfan.

³⁾ Schweregrad mittels der Common Terminology Criteria of Adverse Event (CTCAE) Version 3.0 klassifiziert.

⁴⁾ Dem Vorgehen des pU, das Auftreten eines Todesfalls mit einem UE, das zum Therapieabbruch führte, gleichzusetzen, wird nicht gefolgt. Das Auftreten von UE kann zur Folge haben, dass die Konditionierung nicht begonnen wird oder die OTL-200-Infusion nach der Konditionierung nicht verabreicht wird. Eine Darstellung ebendieser Therapieabbrüche wegen UE erfolgt seitens des pU nicht.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; IDS: Integrierter Datensatz; k. A.: keine Angabe; pU: pharmazeutischer Unternehmer; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Symptome der Grunderkrankung mit CTCAE-Grad ≥ 3 wurden als UE berichtet. Der pU legte keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung mit in die Erfassung der UE eingingen. Insbesondere für die SOC „Erkrankungen des Nervensystems“ kann nicht ausgeschlossen werden, dass Symptome der Endpunktkategorie „Morbidity“ subsumiert wurden.

Unerwünschte Ereignisse

Tabelle 19: UE mit Inzidenz ≥ 20 % des gesamten IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); Sicherheitspopulation¹⁾

IDS UE <i>MedDRA-Systemorganklasse Preferred Term</i>	Studienbeginn bis Konditionierung ²⁾ (Tag -5)		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Erkrankungen des Gastrointestinaltrakts	k. A.	k. A.	20 (100)	9 (100)
Obstipation	k. A.	k. A.	5 (25)	3 (33)
Enteritis	k. A.	k. A.	4 (20)	2 (22)
Stomatitis	k. A.	k. A.	15 (75)	7 (78)
Erbrechen	k. A.	k. A.	9 (45)	2 (22)
Infektionen und parasitäre Erkrankungen	k. A.	k. A.	19 (95)	8 (89)
Konjunktivitis	k. A.	k. A.	5 (25)	1 (11)
Coronavirus-Infektion	k. A.	k. A.	7 (35)	2 (22)
Infektion im Zusammenhang mit einem Medizinprodukt	k. A.	k. A.	8 (40)	0
Ohreninfektion	k. A.	k. A.	7 (35)	0
Gastroenteritis	k. A.	k. A.	12 (60)	3 (33)
Scharlach	k. A.	k. A.	6 (30)	0
Infektion der oberen Atemwege	k. A.	k. A.	12 (60)	6 (67)
Harnwegsinfektion	k. A.	k. A.	3 (15)	4 (44)
Windpocken	k. A.	k. A.	6 (30)	0
Untersuchungen	k. A.	k. A.	18 (90)	7 (78)
Erhöhtes Immunoglobulin E im Blut	k. A.	k. A.	10 (50)	3 (33)
Erniedrigter BMI	k. A.	k. A.	2 (10)	5 (56)
Erhöhtes Serum-Ferritin	k. A.	k. A.	3 (15)	3 (33)
Vitamin D erniedrigt	k. A.	k. A.	8 (40)	2 (22)
Erkrankungen des Blutes und des Lymphsystems	k. A.	k. A.	15 (75)	8 (89)
Febrile Neutropenie	k. A.	k. A.	15 (75)	8 (89)

IDS UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung ²⁾ (Tag -5)		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	k. A.	k. A.	15 (75)	9 (100)
Gangstörung	k. A.	k. A.	10 (50)	7 (78)
Fieber	k. A.	k. A.	10 (50)	3 (33)
Leber- und Gallenerkrankungen	k. A.	k. A.	14 (70)	6 (67)
Gallenblasenpolyp	k. A.	k. A.	4 (20)	2 (22)
Hepatomegalie	k. A.	k. A.	5 (25)	1 (11)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	k. A.	k. A.	12 (60)	5 (56)
Kopfverletzung	k. A.	k. A.	6 (30)	1 (11)
Stoffwechsel- und Ernährungsstörungen	k. A.	k. A.	12 (60)	5 (56)
Eisenmangel	k. A.	k. A.	5 (25)	1 (11)
Metabolische Azidose	k. A.	k. A.	4 (20)	4 (44)
Erkrankungen der Haut und des Unterhautgewebes	k. A.	k. A.	12 (60)	4 (44)
Erythematöser Hautausschlag	k. A.	k. A.	6 (30)	2 (22)
Erkrankungen des Nervensystems	k. A.	k. A.	11 (55)	9 (100)
Aphasie	k. A.	k. A.	4 (20)	6 (67)
Ataxie	k. A.	k. A.	2 (10)	5 (56)
Kognitive Störung	k. A.	k. A.	7 (35)	6 (67)
Motorische Funktionsstörung	k. A.	k. A.	5 (25)	8 (89)
Spastik	k. A.	k. A.	6 (30)	8 (89)
Erkrankungen der Atemwege, des Brustraums und Mediastinums	k. A.	k. A.	10 (50)	4 (44)
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	k. A.	k. A.	9 (45)	7 (78)
Deformität des Fußes	k. A.	k. A.	4 (20)	3 (33)
Osteoporose	k. A.	k. A.	2 (10)	4 (44)
Kongenitale, familiäre und genetische Erkrankungen	k. A.	k. A.	8 (40)	1 (11)
Augenerkrankungen	k. A.	k. A.	8 (40)	3 (33)
Endokrine Erkrankungen	k. A.	k. A.	5 (25)	1 (11)
Erkrankungen der Geschlechtsorgane und der Brustdrüse	k. A.	k. A.	5 (25)	3 (33)
Vorzeitige Menopause	k. A.	k.A.	3 (15)	3 (33)

IDS UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung ²⁾ (Tag -5)		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Gutartige, bösartige und unspezifische Neubildungen (einschl. Zysten und Polypen)	k. A.	k. A.	4 (20)	2 (22)
Erkrankungen der Nieren und Harnwege	k. A.	k. A.	4 (20)	5 (56)
Renaltubuläre Azidose	k. A.	k. A.	1 (10)	5 (56)

¹⁾ Alle Patientinnen und Patienten, die in eine der 4 Studien eingeschlossen wurden und OTL-200 erhalten haben.

²⁾ Zeitraum ab Tag der Unterzeichnung der Einwilligungserklärung bis zum Tag vor Beginn der Konditionierung mit Busulfan.

Abkürzungen: IDS: Integrierter Datensatz; k. A.: keine Angabe; MedDRA: Medical Dictionary for Regulatory Activities; UE: Unerwünschtes Ereignis.

Tabelle 20: Schwere UE (Grad 3 oder höher), die bei ≥ 5 % der Teilpopulation und mindestens 2 Patientinnen und Patienten des gesamten IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029) aufgetreten sind, nach Systemorganklasse und Preferred Term im IDS; Sicherheitspopulation¹⁾

IDS Schwere UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung (Tag -5) ²⁾		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Erkrankungen des Gastrointestinaltrakts	k. A.	k. A.	16 (80)	9 (100)
Dysphagie	k. A.	k. A.	1 (5)	3 (33)
Enteritis	k. A.	k. A.	2 (10)	1 (11)
Stomatitis	k. A.	k. A.	14 (70)	7 (78)
Erbrechen	k. A.	k. A.	5 (25)	1 (11)
Erkrankungen des Blutes und des Lymphsystems	k. A.	k. A.	15 (75)	8 (89)
Atypisches hämolytisches urämisches Syndrom	k. A.	k. A.	2 (10)	⁻³⁾
Febrile Neutropenie	k. A.	k. A.	15 (75)	8 (89)
Neutropenie	k. A.	k. A.	5 (25)	⁻³⁾
Infektionen und parasitäre Erkrankungen	k. A.	k. A.	13 (65)	4 (44)
Clostridium-difficile-Kolitis	k. A.	k. A.	1 (5)	1 (11)

IDS Schwere UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung (Tag -5) ²⁾		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Infektion im Zusammenhang mit einem Medizinprodukt	k. A.	k. A.	5 (25)	_ ³⁾
Gastroenteritis	k. A.	k. A.	3 (15)	_ ³⁾
Pneumonie	k. A.	k. A.	2 (10)	1 (11)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	k. A.	k. A.	11 (55)	7 (78)
Gangstörung	k. A.	k. A.	10 (50)	7 (78)
Erkrankungen des Nervensystems	k. A.	k. A.	11 (55)	9 (100)
Aphasie	k. A.	k. A.	4 (20)	6 (67)
Ataxie	k. A.	k. A.	2 (10)	5 (56)
Kognitive Störung	k. A.	k. A.	7 (35)	6 (67)
Dysarthrie	k. A.	k. A.	1 (5)	4 (44)
Motorische Funktionsstörung	k. A.	k. A.	5 (25)	8 (89)
Spastik	k. A.	k. A.	6 (30)	8 (89)
Krampfanfall	k. A.	k. A.	1 (5)	2 (22)
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen	k. A.	k. A.	6 (30)	6 (67)
Deformität des Fußes	k. A.	k. A.	4 (20)	3 (33)
Deformitäten der Extremitäten	k. A.	k. A.	_ ³⁾	2 (22)
Skoliose	k. A.	k. A.	1(5)	2 (22)
Erkrankungen des Ohrs und des Labyrinths	k. A.	k. A.	4 (20)	_ ³⁾
Hypakusis	k. A.	k. A.	3 (15)	_ ³⁾
Leber- und Gallenerkrankungen	k. A.	k. A.	3 (15)	_ ³⁾
Venöse okklusive Leberkrankheit	k. A.	k. A.	3 (15)	_ ³⁾
Untersuchungen	k. A.	k. A.	3 (15)	1 (11)
Stoffwechsel- und Ernährungsstörungen	k. A.	k. A.	3 (15)	1 (11)
Metabolische Azidose	k. A.	k. A.	2 (10)	1 (11)
Gutartige, bösartige und unspezifische Neubildungen (einschl. Zysten und Polypen)	k. A.	k. A.	2 (10)	_ ³⁾
Erkrankungen der Atemwege, des Brustraums und Mediastinums	k. A.	k. A.	2 (10)	1 (11)
Epistaxis	k. A.	k. A.	1 (5)	1 (11)

IDS Schwere UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung (Tag -5) ²⁾		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Erkrankungen der Haut und des Unterhautgewebes	k. A.	k. A.	2 (10)	1 (11)
Erythematöser Hautausschlag	k. A.	k. A.	2 (10)	1 (11)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	k. A.	k. A.	1 (5)	3 (33)
Gelenksluxation	k. A.	k. A.	1 (5)	3 (33)

¹⁾ Alle Patientinnen und Patienten, die in eine der 4 Studien eingeschlossen wurden und OTL-200 erhalten haben.

²⁾ Zeitraum ab Tag der Unterzeichnung der Einwilligungserklärung bis zum Tag vor Beginn der Konditionierung mit Busulfan.

³⁾ Inzidenz erreicht nicht die angegebene Inzidenzschwelle.

Abkürzungen: IDS: Integrierter Datensatz; MedDRA: Medical Dictionary for Regulatory Activities; k. A.: keine Angabe; UE: Unerwünschtes Ereignis; -: nicht berechenbar.

Tabelle 21: SUE, die bei ≥ 5 % der Teilpopulation und mindestens 2 der Patientinnen und Patienten des gesamten IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029) aufgetreten sind, nach Systemorganklasse und Preferred Term im IDS; Sicherheitspopulation¹⁾

IDS SUE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung (Tag -5) ²⁾		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt ³⁾	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Erkrankungen des Gastrointestinaltrakts	k. A.	k. A.	8 (40)	3 (33)
Enteritis	k. A.	k. A.	2 (10)	– ⁴⁾
Erbrechen	k. A.	k. A.	5 (25)	– ³⁾
Dysphagie	k. A.	k. A.	1 (10)	3 (33)
Infektionen und parasitäre Erkrankungen	k. A.	k. A.	8 (40)	2 (22)
Infektion im Zusammenhang mit einem Medizinprodukt	k. A.	k. A.	2 (10)	– ³⁾
Gastroenteritis	k. A.	k. A.	3 (15)	– ³⁾
Atemwegsinfektion	k. A.	k. A.	2 (10)	– ³⁾
Pneumonie	k. A.	k. A.	1 (5)	1 (11)

IDS SUE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Studienbeginn bis Konditionierung (Tag -5) ²⁾		Ab Beginn Konditionierung (Tag -4) bis Datenschnitt ³⁾	
	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)	OTL-200 Präsympto- matisch N = 20 n (%)	OTL-200 Frühsympto- matisch N = 9 n (%)
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen	k. A.	k. A.	5 (25)	5 (56)
Deformität des Fußes	k. A.	k. A.	4 (20)	3 (33)
Deformitäten der Extremitäten	k. A.	k. A.	– ³⁾	2 (22)
Erkrankungen des Nervensystems	k. A.	k. A.	5 (25)	6 (67)
Aphasie	k. A.	k. A.	2 (10)	3 (33)
Kognitive Störung	k. A.	k. A.	2 (10)	3 (33)
Epilepsie	k. A.	k. A.	1 (5)	3 (33)
Motorische Funktionsstörung	k. A.	k. A.	2 (10)	6 (67)
Spastik	k. A.	k. A.	1 (5)	2 (22)
Krampfanfall	k. A.	k. A.	1 (5)	2 (22)
Erkrankungen der Geschlechtsorgane und der Brustdrüse	k. A.	k. A.	3 (15)	3 (33)
Vorzeitige Menopause	k. A.	k. A.	3 (15)	3 (33)
Leber- und Gallenerkrankungen	k. A.	k. A.	2 (10)	2 (22)
Gallenblasenpolyp	k. A.	k. A.	1 (5)	2 (22)
Stoffwechsel- und Ernährungsstörungen	k. A.	k. A.	1 (5)	1 (11)
Metabolische Azidose	k. A.	k. A.	1 (5)	1 (11)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	k. A.	k. A.	1 (5)	1 (11)
Gelenksluxation	k. A.	k. A.	1 (5)	1 (11)

¹⁾ Alle Patientinnen und Patienten, die in eine der 4 Studien eingeschlossen wurden und OTL-200 erhalten haben.

²⁾ Zeitraum ab Tag der Unterzeichnung der Einwilligungserklärung bis zum Tag vor Beginn der Konditionierung mit Busulfan.

³⁾ Datenschnitte vom 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029).

⁴⁾ Inzidenz erreicht nicht die angegebene Inzidenzschwelle.

Abkürzungen: IDS: Integrierter Datensatz; k. A.: keine Angabe; MedDRA: Medical Dictionary for Regulatory Activities; SUE: Schwerwiegendes unerwünschtes Ereignis; -: nicht berechenbar.

4 Diskussion der Methodik und Ergebnisse

Die vorliegende Nutzenbewertung erfolgt im Rahmen einer Neubewertung nach Befristung. Die Auflagen der Befristung umfassen die Vorlage der finalen Analysen für die pivotale Studie 201222 getrennt nach prä- und frühsymptomatischen Patientinnen und Patienten sowie insbesondere für präsymptomatische Patientinnen und Patienten auch die Vorlage einer Geschwisterkinderanalyse [8, 9, 11].

Die Nutzenbewertung im Vorverfahren stützte sich auf einen integrierten Datensatz (IDS), der die vier Studien 201222, CUP 207394, CUP 206258 und HE 205029 umfasst, welche auch im Zulassungsverfahren durch die EMA Berücksichtigung fanden. Zu diesen Studien wurden mit dem Dossier neue Daten vorgelegt, die zur Nutzenbewertung herangezogen werden.

Mit dem Dossier werden mehrere indirekte Vergleiche vorgelegt, welche verschiedene Mengen der Patientinnen und Patienten aus dem IDS und der Teilnehmenden der natürlichen Verlaufskohorte TIGET-NHx miteinander vergleichen. Diese indirekten Vergleiche werden wegen wesentlicher Unklarheiten nicht zur Nutzenbewertung herangezogen. Aus Transparenzgründen erfolgt jedoch im Anhang eine Darstellung einer der indirekten Vergleiche, der aus dem IDS und der TIGET-NHx ein Matchingset auf Basis von Geschwistern in den beiden Studien erstellt (siehe Kapitel 4.1).

Im Rahmen der Neubewertung nach Befristung reicht der pU zudem einen Studienbericht für die Studie 205756 ein, eine einarmige Studie zur kryokonservierten Formulierung von OTL-200, zu der für die vorangegangene Bewertung [10] noch keine Ergebnisse vorgelegt wurden. Auch diese Studie wurde im selben Studienzentrum durchgeführt und wird als hinreichend vergleichbar mit den anderen Studien des IDS angesehen, sodass eine Aufnahme der Cryo-Studie 205756 in das IDS grundsätzlich gerechtfertigt erscheint. Gemäß EPAR liegen zudem hinreichend Daten vor, die eine Vergleichbarkeit der frischen und der kryokonservierten Formulierung belegen [4]. Mit dem Dossier wurden Ergebnisse zu der Studie 205756 mit einem Datenschnitt vom 22.08.2022 vorgelegt. Die Studie wurde jedoch am 10.03.2026 vorzeitig beendet, welches auch das Datum des finalen Datenschnitts darstellt. Zu diesem Datenschnitt wurden für die Studie 205756 jedoch keine Ergebnisse vorgelegt. Für eine in Modul 4 berichtete gepoolte Analyse des IDS mit der Studie 205756 liegen widersprüchliche Angaben zum Datenschnitt vor, sodass die Aktualität der Daten nicht bewertet werden kann. Eine ergebnisgesteuerte Berichterstattung kann daher für die Studie 205756 nicht ausgeschlossen werden und diesbezügliche Ergebnisse werden daher in der vorliegenden Nutzenbewertung nicht berücksichtigt (siehe auch Tabelle 1).

Zudem reichte der pU Studienunterlagen zur Studie LongTERM-MLD (OTL-200-10, EUPAS48374) ein. Die laufende prospektive Langzeitnachsbeobachtungsstudie wurde von der EMA im Oktober 2020 angefordert und startete im Dezember 2022. Sie dient der Beurteilung der langfristigen klinischen Wirksamkeit und Sicherheit auf Basis eines Registers. Im Zuge der Studie werden u. a. auch gepoolte Ergebnisse des IDS und der Studie 205756 berichtet, die jedoch auf Datenschnitten von 2022 und 2024 basieren, welche älter sind als die Daten, die für das IDS mit Datenschnitten vom Mai und Juni 2025 zur Nutzenbewertung herangezogen werden. Somit liegen noch keine Daten vor, die im Rahmen der eigentlichen LongTERM-Studie erhoben wurden. Folglich wird die Studie LongTERM-MLD nicht zur Nutzenbewertung herangezogen. Gemäß pU ist das geplante Ende der Studie für das Jahr 2042 prognostiziert, gemäß EMA für den 29. März 2041 [15].

Die vorliegende Nutzenbewertung stützt sich demnach auf den integrierten Datensatz (IDS), der die vier Studien 201222, CUP 207394, CUP 206258 und HE 205029 umfasst. Ergebnisse des indirekten Vergleichs „Geschwisterkinderanalyse IDS vs. TIGET-NHx“ sind im Anhang dargestellt.

4.1 Design und Methodik der Studie

Studiendesign der Studien im IDS

Die Studie 201222 war eine nicht-randomisierte, offene, einarmige, monozentrische Phase-I/II-Studie, in der die Wirksamkeit und Sicherheit der Studienmedikation OTL-200 bei Kindern mit LI- oder EJ-Form der MLD untersucht wird. In einem Studienzentrum in Mailand (Italien) wurde am 09.04.2010 die erste Testperson eingeschlossen. Das Studienende (definiert als das Datum der Datenbanksperrung nach der letzten Visite der letzten Studienperson) war am 19.09.2025. Mit Protokoll-Amendment 6 vom 10.12.2013 wurde eingeführt, dass nur Patientinnen und Patienten eingeschlossen werden, bei denen, bezogen auf die MLD-Formen, 2 der 3 im Folgenden aufgelisteten Kriterien erfüllt wurden: a) Alter des älteren Geschwisterkinds beim Auftreten der Symptome (LI: ≤ 30 Monate; EJ: 30 Monate – 6 Jahre), b) Genotyp (LI: 2 Null (0)-mutierte ARSA-Allele; EJ: 1 Null (0)- und 1 R-mutiertes ARSA-Allel) und c) periphere Neuropathie mittels Elektroneurographie (ENG). Für die bereits vor dem Zeitpunkt der Protokolländerung eingeschlossenen Patientinnen und Patienten ist unklar, ob sie alle 3 Kriterien erfüllen mussten. Für fröhsymptomatische Kinder wurde zusätzlich das Einschlusskriterium, dass die Symptome der Erkrankung nicht länger als 6 Monate vorliegen durften, durch die folgenden Kriterien ersetzt: a) IQ ≥ 70 Punkte und b) Fähigkeit, unabhängig mindestens 10 Schritte gehen zu können. Bis Protokoll-Amendment 6 wurden laut Modul 4 2 fröhsymptomatische Patientinnen und Patienten, gemäß der ursprünglichen Definition, in die Studie aufgenommen. Es ist unbekannt, inwiefern sich die Teilpopulationen, die vor und nach der Protokolländerung eingeschlossen wurden, voneinander unterscheiden.

Die Studie war in 4 Phasen unterteilt: a) Screening-Phase, b) Baseline-Phase, c) eine 36-tägige Behandlungsphase, die die Gewinnung hämatopoetischer Stammzellen, Apherese, myeloablative Konditionierung mit Busulfan sowie die Verabreichung der Studienmedikation umfasst, und d) eine mindestens 8-jährige Nachbeobachtungsphase.

Die Studie 201222 wurde im IDS mit Daten aus 2 Härtefallprogrammen CUP 207394 und CUP 206258 und der Hospital Exemption HE 205029 ergänzt. Bei Studie CUP 207394 handelt es sich um einen Fallbericht eines am 23.04.2013 eingeschlossenen fröhsymptomatischen Patienten mit EJ-MLD, dem die Studienmedikation OTL-200 im Rahmen eines Härtefallprogramms zur Verfügung gestellt wurde. Zu diesem Zeitpunkt wurden keine EJ-Patientinnen und -Patienten in die Studie 201222 rekrutiert. Studie CUP 206258 ist eine Fallserie von Patientinnen und Patienten mit LI- oder EJ-Form mit frühzeitiger klinischer Manifestation, die präsymptomatisch waren und ab dem 16.01.2017 als Härtefall eingeschlossen wurden. Mit Protokoll-Amendment 1 vom 12.07.2017 wurde eingeführt, dass, neben LI-Patientinnen und -Patienten, auch Kindern mit der intermediären oder EJ-Form der Erkrankung die Studienmedikation verabreicht werden durfte. Für die Studie HE 205029 liegt ein gemeinsamer Studienbericht mit Studie CUP 206258 vor. Es handelt sich bei der Studie HE 205029 um eine Fallserie von Patientinnen und Patienten mit LI-Form, die mit der Studienmedikation OTL-200 ab dem 29.12.2015 im Rahmen einer Hospital Exemption behandelt wurden. Alle 4 Studien zeichnet ein vergleichbares Studiendesign und eine Durchführung in einem ausgewählten Zentrum in Mailand (Italien) aus. Angesichts vergleichbarer Studienbedingungen und hinreichend ähnlicher Studienpopulationen, aber auch aufgrund der kleinen Fallzahl in der Hauptstudie 201222, erscheint eine gemeinsame Auswertung als integrierter Datensatz adäquat.

Das Anwendungsgebiet gemäß Fachinformation unterscheidet folgende Populationen:

- a) Kinder mit LI- oder EJ-MLD, ohne klinische Manifestation der Erkrankung (= präsymptomatisch).

- b) Kinder mit EJ-MLD mit frühzeitiger klinischer Manifestation der Erkrankung, die jedoch noch selbstständig gehen können, vor dem Einsetzen einer kognitiven Verschlechterung (= frühsymptomatisch).

Eine getrennte Ergebnisdarstellung für prä- und frühsymptomatische Patientinnen und Patienten entspricht damit der Unterteilung des Anwendungsgebiets. Für die Nutzenbewertung werden daher die Wirksamkeits- und Sicherheitsergebnisse des IDS getrennt für die beiden Teilpopulationen herangezogen.

In der vorherigen Nutzenbewertung von OTL-200 [6, 7, 10, 28] lagen für die Studien des IDS Datenschnitte zwischen Januar und Dezember des Jahres 2018 vor. Für die aktuelle Nutzenbewertung liegen jeweils die finalen Datenschnitte der Studien vor, die zwischen Mai und Juni 2025 liegen. Für die Studien des IDS beträgt die zusätzliche Studiennachbeobachtungsdauer somit zwischen 6 und 7 Jahren. Die Rekrutierung in der pivotalen Studie war zum Zeitpunkt des letzten Datenschnitts (30.03.2018) bereits abgeschlossen. Die aktuelle Stichprobengröße des IDS entspricht der Stichprobengröße in der Vorbewertung (ITT-Population: N = 29). Es ist zu beachten, dass die in Modul 4 vom pU angegebenen finalen Datenschnitte nicht anhand der vorgelegten Studienberichte verifiziert werden konnten. Den Studienberichten der jeweiligen Studien des IDS können keine konkreten Datenschnitte entnommen werden.

Für behandelte Patientinnen und Patienten der pivotalen Studie sowie für die Studienperson der Studie CUP 207394 (72 % des IDS) war eine Nachbeobachtungszeit von mindestens 8 Jahren vorgesehen, für Teilnehmende der Studien CUP 206258 und HE 205029 (28 % des IDS) von bis zu 8 Jahren. Daraus ergeben sich unterschiedliche geplante Nachbeobachtungsdauern zwischen den Studien des IDS und gemäß Planung sollte die Erhebung von Endpunkten ab dem 8. Jahr bei 28 % des IDS eingestellt worden sein. In den beiden erstgenannten Studien mit 72 % des IDS war nach Jahr 8 zwar eine Nachbeobachtung möglich, jedoch lediglich eine jährliche Endpunkterfassung vorgesehen. Deshalb ist bspw. unklar, ob die Sicherheitserfassung ab Jahr 8 vollständig war. Für alle Teilnehmenden der IDS-Studien besteht seit Dezember 2022 die Möglichkeit, in die Langzeitnachbeobachtungsstudie OTL-200-10 überzugehen.

Studienpopulation

Es wurden insgesamt 31 Patientinnen und Patienten in die Studien 201222, CUP 207394, CUP 206258 und HE 205029 eingeschlossen. Aufgrund eines Widerrufs einer erteilten Einwilligungserklärung vor Behandlungsbeginn in der Studie 201222 umfasst die Gesamtpopulation für die gemeinsame Analyse der Studien 201222, CUP 207394, CUP 206258 und HE 205029 ($\triangleq$ IDS) 30 Patientinnen und Patienten. Eine Person ist anlässlich klinischer Kriterien zu Baseline aus der Studie 201222 ausgeschieden und erhielt die OTL-200-Infusion nicht. Die ITT- und Sicherheitspopulation bilden insgesamt 29 Patientinnen und Patienten; davon waren 20 prä- und 9 frühsymptomatisch. Insgesamt 12 prä- und 8 frühsymptomatische Kinder entstammen der pivotalen Studie, 1 frühsymptomatisches Kind dem Compassionate-Use-Program (CUP) 207394, 5 präsymptomatische Kinder dem CUP 206258 sowie 3 präsymptomatische Kinder dem Hospital-Exemption (HE)-Program 205029. Es muss berücksichtigt werden, dass die vom pU definierte ITT-Population nicht alle Personen berücksichtigt, die in eine der Studien des IDS eingeschlossen wurden und die möglicherweise die Vorbereitungen auf die OTL-200-Behandlung schon begonnen hatten, aber nicht mit OTL-200 behandelt wurden (siehe auch Kapitel 2.4). Im gesamten IDS gab es 4 Studienabbrüche: 3 Kinder verstarben nach der Gentherapie (präsymptomatisch: n = 1; frühsymptomatisch: n = 2) und 1 weiteres Kind konnte im Verlauf der Studie nicht weiter nachbeobachtet werden (Lost to Follow-up; unklar, ob prä- oder frühsymptomatisch).

Präsymptomatische MLD

In der Teilpopulation der präsymptomatischen Patientinnen und Patienten betrug das mediane Alter zum Zeitpunkt der Diagnosestellung knapp 9 Monate (min; max: 0; 61), zu Studieneinschluss 12 Monate (min; max: 7; 62). Insgesamt waren die Patientinnen und Patienten überwiegend männlich (65 %). Ob alle zu Studieneinschluss präsymptomatischen Patientinnen und Patienten zum Zeitpunkt der Gentherapie noch präsymptomatisch waren, kann den Studienunterlagen nicht entnommen werden. Bei 15 Kindern (75 %) war LI-MLD diagnostiziert worden, bei 5 (25 %) EJ-MLD. Für die Beurteilung grobmotorischer Funktionen wurde ein mittlerer GMFM-Wert von 59 % (SD: 26) zu Baseline berichtet, wobei das junge Alter der Patientinnen und Patienten zu Studieneinschluss berücksichtigt werden muss, da der GMFM sich an einem gesunden und motorisch normal entwickelten 5-jährigen Kind orientiert. 4 Patientinnen und Patienten (20 %) waren, dem GMFC-MLD zufolge, zu Baseline ohne Hilfe gehfähig (GMFC-MLD-Level 0). Es wurden für einen Großteil der Teilpopulation (80 %) jedoch keine GMFC-MLD-Werte dargestellt. Fehlende Werte können zum Teil darin begründet sein, dass der Einsatz des Messinstruments erst ab einem Alter von 18 Monaten empfohlen wird. Die Beobachtungsdauer nach der OTL-200-Infusion betrug im Median 10 Jahre (min; max: 1; 15).

Frühsymptomatische MLD

In der Teilpopulation der frühsymptomatischen Patientinnen und Patienten betrug das mediane Alter zum Zeitpunkt der Diagnosestellung ca. 68 Monate (min; max: 20; 132), zu Studieneinschluss knapp 70 Monate (min; max: 22; 136). Die Patientinnen und Patienten waren überwiegend weiblich (67 %), bei 8 Kindern (89 %) wurde eine EJ-MLD diagnostiziert, bei einem Kind (11 %) LI-MLD. Der mittlere GMFM-Wert zu Baseline betrug 85 % (SD: 10) und alle Patientinnen und Patienten waren dem GMFC-MLD zufolge zu Baseline ohne Hilfe (22 %, Level 0) oder mit Hilfe (78 %, Level 1) gehfähig. Die höheren GMFM-Werte zu Baseline in der frühsymptomatischen Teilpopulation (im Vergleich zu den präsymptomatischen Patientinnen und Patienten) könnten darin begründet sein, dass die frühsymptomatischen Kinder zu Studieneinschluss mit einem medianen Alter von knapp 6 Jahren das Referenzalter des GMFM von 5 Jahren knapp überschreiten, während die präsymptomatischen Kinder dieses mit einem medianen Alter von etwa 1 Jahr deutlich unterschreiten. Die Beobachtungsdauer nach der OTL-200-Infusion betrug im Median 11 Jahre (min; max: 1; 14).

Studienmedikation

Alle 29 behandelten Patientinnen und Patienten erhielten vor der OTL-200-Infusion eine Konditionierung unter Verwendung von Busulfan.

- **Präsymptomatische MLD**

Die eine Hälfte der präsymptomatischen Patientinnen und Patienten (n = 10) erhielt eine sub-myeloablative Konditionierung (SMAC) mit 14 Dosen Busulfan als 2-stündige intravenöse Infusion, die andere Hälfte (n = 10) eine myeloablative Konditionierung (MAC) mit insgesamt 4 Dosen Busulfan als 3-stündige intravenöse Infusion. Die Studienmedikation wurde 1-mal mit einer medianen Zelldosis von $12,5 \times 10^6$ CD34+-Zellen/kg (min; max: 4,2; 25,9) als intravenöse Infusion innerhalb von 10 bis 20 Minuten verabreicht.

- **Frühsymptomatische MLD**

In der frühsymptomatischen Teilpopulation erhielten 3 Kinder (33 %) eine SMAC und 6 (67 %) eine MAC. Die mediane Zelldosis bei Verabreichung der Studienmedikation betrug bei den frühsymptomatischen Kindern $7,1 \times 10^6$ CD34+-Zellen/kg (min; max: 3,8; 13,1).

Die Angaben zur Mindestdosis entsprechen der Fachinformation, der zufolge die empfohlene Mindestdosis 3×10^6 CD34+-Zellen/kg beträgt. Gemäß Fachinformation sollte das maximal zu verabreichende Volumen der Studienmedikation weniger als 20 % des geschätzten

Plasmavolumens der Patientin oder des Patienten betragen. Es ist unklar, in welchem Ausmaß der Empfehlung in den der Nutzenbewertung zugrundeliegenden Studien gefolgt wurde.

Angaben zu Vor- und Begleittherapie sind für den IDS nicht berichtet. Für die Studie 201222 können, dessen ungeachtet, nur unvollständige Angaben zur Begleitbehandlung gemacht werden, da seitens pU beispielhaft aufgelistete Begleitmedikamente, die (z. B. Antiemetika, Analgetika und Antiepileptika) routinemäßig im Studienzentrum eingesetzt werden, ausschließlich in solchen Fällen im Fallberichtsbogen erfasst werden sollten, in denen sie wegen einer Begleiterkrankung oder UE zum Einsatz kamen.

Indirekter Vergleich

Im Zuge des vorherigen Nutzenbewertungsverfahrens zu OTL-200 wurde eine Analyse einer Matchingpopulation aus Geschwisterkindern des IDS und der natürlichen Verlaufsstudie TIGET-NHx zur Ableitung des Zusatznutzens herangezogen. Eine Geschwisterkinderanalyse des IDS vs. TIGET-NHx zu präsymptomatischen Patientinnen und Patienten wurde auch im Zuge der Befristungsaufgaben angefordert. [8, 9, 11]

Der pU reicht insgesamt 4 neue indirekte Vergleiche auf Basis von neuen Daten ein, in denen mit OTL-200 behandelte Patientinnen und Patienten aus den Studien des IDS und/oder der Cryo-Studie 205756 mit historischen Kontrollen aus der TIGET-NHx-Studie gegenübergestellt werden:

- A Geschwisterkinderanalyse pivotale Studie 201222 vs. TIGET-NHx (MSAS)
- B Frühsymptomatische Personen des IDS vs. TIGET-NHx (ITT-Population)
- C **Geschwisterkinderanalyse IDS vs. TIGET-NHx (MSAS)**
- D Geschwisterkinderanalyse IDS + Cryo-Studie 205756 vs. TIGET-NHx (MSAS)

Der indirekte Vergleich C zur Geschwisterkinderanalyse IDS vs. TIGET-NHx mit 94 % präsymptomatischen Kindern im IDS-Arm und ausschließlich mindestens frühsymptomatischen Kindern im TIGET-NHx-Arm kommt der Befristungsaufgabe am nächsten und wird vom pU in Modul 4 zur Ableitung des Zusatznutzens herangezogen. Da für die Nutzenbewertung das IDS als maßgebliche Datengrundlage angesehen wird (siehe u. a. Kapitel 4.1), wird auch in der vorliegenden Nutzenbewertung primär zu diesem indirekten Vergleich C ausgeführt (siehe auch Kapitel 2.6). Der indirekte Vergleich B, welcher eine Teilpopulation des IDS, frühsymptomatische Studienteilnehmende, umfasst, wird nicht zur Nutzenbewertung herangezogen, da umfangreiche Limitationen und eine fehlende Strukturgleichheit der Vergleichsarme vorliegen (siehe Kapitel 2.6.3). Der Vergleich D, der die Studie 205756 (Cryo) zur kryokonservierten Formulierung von OTL-200 einbezieht, wird aufgrund der Hinweise auf eine ergebnisgesteuerte Berichterstattung in Bezug auf die Cryo-Studie nicht für die Nutzenbewertung berücksichtigt (siehe auch Kapitel 4).

In Modul 4 führt der pU aus, dass es sich bei der Geschwisterkinderanalyse IDS vs. TIGET-NHx (indirekter Vergleich C) um einen geeigneten Vergleich handeln würde, da Geschwister mit dem gleichen Genotyp und der gleichen familiären und sozialen Umgebung die geringste Variabilität in den Ausgangsbedingungen und im Krankheitsverlauf zeigen sollten. Diese Annahme kann nachvollzogen werden und vor dem Hintergrund des fortschreitenden natürlichen Verlaufs der MLD, bei der Spontanremissionen nicht zu erwarten sind, erscheint ein Vergleich dieser Art grundsätzlich sinnvoll. Aus den Einreichungen des pU ergeben sich jedoch umfangreiche Unklarheiten und Limitationen zu dem durchgeführten indirekten Vergleich, woraus wesentliche Unsicherheiten hinsichtlich der zentralen Annahmen der kausalen Inferenz (Positivität, Überlappung und Balance) resultieren. So scheint bspw. nicht sichergestellt, dass die gematchte Population ausschließlich tatsächlich verwandte Geschwisterpaare enthielt, da kein 1:1-Matching erfolgen sollte.

Unklarheiten und/oder Limitationen bestehen zusammengefasst aus folgenden Punkten (nicht abschließend), die in Kapitel 2.6 weiter ausgeführt sind:

- Fehlende systematische Confounderidentifikation und vollständige Confunderdarstellung zu Baseline/Matchingbaseline vor und nach Matching.
- Fehlende Informationen zum Ausgangsset von allen Geschwisterkindern beider Vergleichsarme (IDS und TIGET-NHx; bspw. Stichprobengröße, Baselinecharakteristika, etc.) und welche Kinder warum nicht gematcht werden konnten.
- Als inadäquat eingeschätztes Matchingkriterium anhand der Altersspanne der mit OTL-200 behandelten Studienteilnehmenden 2 Jahre nach Baseline (statt zu Baseline), wodurch u. a. Personen des IDS, die vor der Visite zu Jahr 2 verstarben (n = 3 im IDS) oder vor Behandlung mit OTL-200 die Studie abbrachen (n = 2) per se vom Matching ausgeschlossen worden zu sein scheinen, unabhängig davon, ob sie Geschwisterkinder in der natürlichen Verlaufskohorte hatten (Selektions- und Immortal Time Bias); Selektions- und Immortal Time Bias liegen für den TIGET-NHx-Arm durch das Kriterium der Altersspanne möglicherweise auch vor.
- Fehlende eindeutige Baselinedefinition, insbesondere für die Geschwisterkinder der Studie TIGET-NHx und die dazugehörigen Baselinecharakteristika nach Matching (siehe auch den ersten Punkt zu Confoundern); der Gesundheitszustand der gematchten Kinder zu Matchingbaseline ist somit in Teilen unklar.
- Vorliegende Informationen stellen insbesondere Positivität und Balance in Frage:
 - Einschluss von Kindern in die Studie TIGET-NHx (Start im Jahr 2004), die für die Teilnahme an den Studien des IDS (Start im Jahr 2010) nicht geeignet waren, sodass auch nach dem Jahr 2010 Kinder mit schwereren kognitiven und motorischen Beeinträchtigungen nicht in die Studien des IDS, sondern in die Studie TIGET-NHx aufgenommen wurden (Selektionsbias).
 - Mindestens 73 % im gematchten Kontrollarm entsprechen bereits zu Screening/Studieneinschluss nicht dem zugelassenen Anwendungsgebiet von OTL-200, da bei diesen eine frühsymptomatische LI-MLD vorliegt; bezüglich des Symptomstatus zeigen sich bereits zu Screening/Studieneinschluss relevante Unterschiede zwischen den Armen und zu Matchingbaseline liegen hierzu keine Informationen vor.
 - Manche der vorliegenden Patientencharakteristika deuten auf relevante Unterschiede zwischen den Armen hin (bspw. Symptomatik zu Screening/Studieneinschluss, Geschlecht und GMFM-Wert).
 - Die aufgeführten Punkte könnten nur teilweise durch retrospektive Daten adressiert werden (siehe jedoch Folgepunkt).
- Fehlende Informationen zu (insbesondere Umfang und Qualität der) verwendeten retrospektiven Daten für die Geschwisterkinder.
- Unklarheiten zum Vorliegen fehlender Daten im TIGET-NHx-Arm zu Jahr 2, wenn beim Matching der früheste vollständige Datensatz genutzt werden sollte.
- Fehlende Informationen zum Umgang mit fehlenden Daten im Zuge der Geschwisterkinderanalyse.
- Keine Informationen zu erfolgten supportiven Interventionen nach Matchingbaseline, insbesondere (aber nicht ausschließlich) zum TIGET-NHx-Arm.
- Weitere endpunktspezifische Unklarheiten (siehe Anhang), bspw. zu retrospektiver Erhebung und widersprüchlichen Rücklaufquoten.

Grundsätzlich wird ein Vergleich von Geschwisterkindern aus dem IDS mit geeigneten natürlichen bzw. nur supportiv behandelten Kontrollen als sinnvoll erachtet. Wie jedoch dargelegt, liegen auch abgesehen von der fraglichen Gegebenheit der Positivität zu dem aktuell vorgelegten indirekten Vergleich umfangreiche Unklarheiten und Limitationen vor, sodass die Eignung der gematchten Vergleichspopulation und der durchgeführten Geschwisterkinderanalyse im Ganzen vorliegend nicht festgestellt werden kann. Eine Darstellung der Ergebnisse des indirekten Vergleichs C (Geschwisterkinderanalyse IDS vs. TIGET-NHx) erfolgt aus Transparenzgründen sowie aufgrund des sehr deutlichen Ausmaßes des Effekts und statistischen Signifikanzniveaus einzelner Ergebnisse dennoch im Anhang.

4.2 Zulassungsstatus und Zulassungspopulation

Zulassungsstatus und Anwendungsgebiet

Atidarsagen autotemcel/OTL-200 (Libmeldy®) ist, laut Fachinformation, indiziert zur Behandlung von metachromatischer Leukodystrophie (MLD) bei a) Kindern mit einer im späten Säuglings- oder frühen Kindesalter auftretenden Form ohne klinische Manifestation der Erkrankung und b) Kindern mit der im frühen Kindesalter auftretenden Form mit einer frühzeitigen klinischen Manifestation der Erkrankung, die jedoch noch selbstständig gehen können, und vor dem Einsetzen einer kognitiven Verschlechterung.

In die für die Nutzenbewertung vorliegenden Studien der integrierten Analyse (Studien 201222, CUP 207394, CUP 206258, HE 205029 $\triangleq$ IDS) wurden Patientinnen und Patienten mit der spät-infantilen Form (LI) ohne eine klinische Manifestation der Erkrankung (präsymptomatisch) oder mit der frühjuvenilen Form (EJ), sowohl ohne klinische Manifestation als auch mit ersten Symptomen (prä- und frühsymptomatisch), eingeschlossen. Da bei einem präsymptomatischen Kind, das in die Studie 201222 eingeschlossen wurde, 2 mutierte Allele mit Restfunktion (R) der ARSA vorlagen, wurde bei ihm eine intermediäre Form der Erkrankung in Betracht gezogen, die zwischen der gängigen LI- und EJ-Form verortet wird. Der pU führt hierzu aus, dass die intermediäre MLD vorliegt, insofern erste Symptome beim älteren Geschwisterkind in einem Alter von ≤ 6 Jahren auftraten und es nicht möglich ist, das Geschwisterkind eindeutig der LI- oder EJ-MLD zuzuordnen. Für die Analysen wurde die intermediäre Form als EJ-MLD klassifiziert. Es wird für das vorliegende Anwendungsgebiet davon ausgegangen, dass die LI-, EJ- und intermediäre MLD umfasst sind.

Mit Protokoll-Amendment 6 der Studie 201222 vom 10.12.2013 wurde eingeführt, dass die Eignung von frühsymptomatischen Patientinnen und Patienten mit EJ-MLD für die Behandlung mit OTL-200 unmittelbar vor der Behandlung ein weiteres Mal überprüft werden soll. Es war in der Folge, im Fall einer bedeutsamen Krankheitsprogression vor Behandlungsbeginn, ein vorzeitiger Studienabbruch geplant. Aus den Unterlagen des pU geht nicht hervor, anhand welcher Kriterien über das Vorliegen einer bedeutsamen Krankheitsprogression entschieden wurde. Es wird vermutet, dass der Abbruch bei einer kognitiven Verschlechterung (IQ < 70 Punkte) und dem Verlust der Fähigkeit, unabhängig mindestens 10 Schritte gehen zu können, erfolgte (siehe Definition frühsymptomatisch, Tabelle 2).

Im Rahmen des Stellungnahmeverfahrens der Vorbewertung [12] beschreibt der pU, dass 2 der 3 verstorbenen Patientinnen und Patienten des IDS nicht vom aktuellen Label umfasst seien. Dies kann nicht nachvollzogen werden. Beide dieser Patientinnen kamen gemäß Angaben des pU nach der Behandlung in eine rapide Progressionsphase und sind an einer mit der Erkrankung verbundenen Dysphagie 8 bzw. 15 Monate nach Behandlung verstorben. Laut Fachinformation muss zum Zeitpunkt der Anwendung des Medikaments der fachinformationskonforme Gesundheitszustand vorliegen. Somit darf eine Progression zum Zeitpunkt der Behandlung nicht über die Einschlusskriterien hinaus fortgeschritten sein. Inwiefern eine rasche Progression nach der Behandlung mit OTL-200 nicht dieser Vorgabe

entspricht, kann nicht nachvollzogen werden. Wenn sie zum Zeitpunkt der Verabreichung der OTL-200-Infusion weiterhin fröhsymptomatisch waren, fallen die 2 Patientinnen in das Anwendungsgebiet von Atidarsagen autotemcel/OTL-200. Weiterhin soll die Behandlung mit OTL-200 gemäß Fachinformation durchgeführt werden, bevor die Erkrankung in ihre schnell fortschreitende Phase gelangt. Anhand welcher Kriterien in den Studien des IDS festgestellt wurde, ob Patientinnen und Patienten in die schnell fortschreitende Phase übergegangen sind, ist nicht ersichtlich. Mit Protokollversion 6 (10.12.2013) der pivotalen Studie wurde zudem die Anforderung eingeföhrt, dass die Eignung von fröhsymptomatischen EJ-Patientinnen und -Patienten unmittelbar vor der Behandlung durch das Prüfpersonal erneut zu überprüfen ist und Personen aus der Studie (vor der Behandlung) auszuschließen sind, wenn es zwischen Studieneinschluss und geplanter Behandlung zu einem schnellen und schweren Fortschreiten der Erkrankung kam. Wären die beiden Patientinnen also zum Zeitpunkt der Gentherapie nicht mehr fröhsymptomatisch gewesen, hätten sie aus der Studie ausgeschlossen werden müssen und hätten die Therapie nicht erhalten dürfen, sofern sie nach dieser Protokolländerung in die Studie eingeschlossen wurden.

Der pU erwähnt in der Stellungnahme [12] zudem 2 weitere Off-Label-Studienteilnehmende, die aufgrund ihres GMFC-MLD-Levels und/oder IQ zum Zeitpunkt der Behandlung mit OTL-200 nicht vom Anwendungsgebiet umfasst gewesen seien. Auch dies kann nicht nachvollzogen werden, da dafür keine Nachweise vorgelegt wurden. Gemäß Studienunterlagen wurde für keine teilnehmende Person ein GMFC-MLD-Level von größer als 1 berichtet und für den IQ liegen keine validen Baseline-Werte vor.

Eine Person in der fröhsymptomatischen Teilpopulation wurde zudem als LI-MLD klassifiziert und entspricht somit nicht dem Anwendungsgebiet. Dem Stellungnahmeverfahren des Vorverfahrens [6, 12] kann entnommen werden, dass diese Person von den bereits genannten 4 Fällen umfasst sein müsste.

In der Gesamtschau könnten bis zu 4 Personen des IDS (14 %) nicht der Zulassungspopulation entsprechen, nachvollziehbare Belege dazu fehlen jedoch.

Sensitivitätsanalysen für die ITT-Population des IDS getrennt nach prä- und fröhsymptomatisch ohne diese potentiellen Off-Label-Patientinnen und -Patienten wurden für die vorliegende Bewertung nicht eingereicht.

Übertragbarkeit auf deutschen Versorgungskontext

Es bestehen, in Anbetracht des monozentrischen Designs der Studien, mit Durchführung an einer spezialisierten Klinik, Unsicherheiten hinsichtlich der Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext.

Unklar erscheint u. a. auch, ob in der Versorgung vergleichbare Zeitabstände zwischen Knochenmarkentnahme (mit darauffolgender Herstellung der OTL-200-Infusion sowie Konditionierung mit Busulfan) und der Behandlung mit der Gentherapie vorliegen. Dies ist insbesondere deswegen relevant, da ein Progress über frühzeitige klinische Manifestationen hinaus vor Verabreichung der OTL-200-Infusion dazu führen würde, dass die Behandlung zu diesem Zeitpunkt nicht mehr der Zulassung entspräche. Bei einer Person aus der Studie 201222 führte eine rasche Krankheitsprogression zwischen dem Screening und einer erneuten Begutachtung zum Zeitpunkt der Baseline-Erhebung (Tag -5) zu einem vorzeitigen Studienabbruch, da die Person aufgrund der Progression nicht mehr für die Behandlung geeignet war. Dies unterstreicht, dass eine zeitnahe Behandlung nach den notwendigen Maßnahmen zur Herstellung von OTL-200 (Knochenmarkentnahme inklusive Stammzellreserve) wichtig sein kann, um eine erfolgreiche Behandlung zu gewährleisten.

4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit

Das Verzerrungspotential der 4 Studien wird angesichts des unkontrollierten Designs auf Studien- und Endpunktebene als hoch eingeschätzt.

4.4 Mortalität

In der Endpunktkategorie „Mortalität“ wurden in der vorliegenden Nutzenbewertung Ergebnisse zum Gesamtüberleben berücksichtigt.

Es zeigte sich zu den jeweiligen Datenschnitten in den Studien 201222 (30.06.2025), CUP 207394 (30.06.2025), CUP 206258 (26.05.2025) und HE 205029 (26.05.2025) ($\triangle$ IDS), dass insgesamt 3 von 29 behandelten Patientinnen und Patienten (10 %) mit einer EJ-MLD verstorben sind.

- Präsymptomatische MLD

Eine dieser Personen (männlich) entstammt der Studie CUP 206258, verstarb zu Monat 14 nach der OTL-200-Infusion aufgrund eines ischämischen Schlaganfalls. Die mediane Beobachtungsdauer ab Gabe der Gentherapie lag bei etwa 10 Jahren [Min: 1,1 Jahre; Max: 15,2 Jahre].

- Frühsymptomatische MLD

Die beiden anderen Personen (beide weiblich) entstammen der Studie 201222, verstarben zu Monat 8 und Monat 15 nach OTL-200 aufgrund einer Dysphagie, infolge des Fortschreitens ihrer Erkrankung. Die Kinder wurden auf Wunsch ihrer Eltern nicht mit einer perkutanen endoskopischen Gastrostomie-Sonde ernährt. Die mediane Beobachtungsdauer ab Gabe der Gentherapie lag bei knapp 11 Jahren [Min: 0,6 Jahre; Max: 13,9 Jahre].

Im Vergleich zur Vorbewertung [10] werden trotz einer längeren Nachbeobachtungszeit keine weiteren Todesfälle berichtet.

Es ist zu beachten, dass für die Studie HE 205029 (und CUP 206258) eine Nachbeobachtung nach Jahr 8 nicht vorgesehen war, zwischen Studienbeginn (Einschluss der ersten Testperson: 29.12.2015) und finalem Datenschnitt gemäß Modul 4 (26.05.2025) jedoch knapp 10 Jahre liegen. Es ist unklar, inwiefern nach Jahr 8 noch eine vollständige Erfassung der Mortalität gegeben war. Dies betrifft maximal N = 3 Personen. Für die Studie CUP 206258 (N = 5) lagen nur 8 Jahre und 4 Monate zwischen Studienbeginn (Einschluss der ersten Testperson: 16.01.2017) und finalem Datenschnitt (26.05.2025), sodass die Erfassung der Mortalität spätestens zu diesem Nachbeobachtungszeitraum für den Datenschnitt endete.

Der Effekt von OTL-200 auf die Mortalität kann auf Grundlage der vorgelegten einarmigen Studiendaten mit einhergehendem hohen Verzerrungspotential nicht abschließend beurteilt werden.

4.5 Morbidität

In der Kategorie „Morbidität“ werden in der Nutzenbewertung Ergebnisse des Endpunkts „GMFM“ berücksichtigt. Die Ergebnisse der Endpunkte „GMFC-MLD“ und „Überleben ohne schwere motorische Beeinträchtigung“ werden vor dem Hintergrund der in Kapitel 2.3 jeweils beschriebenen Limitationen nicht berücksichtigt. Für den „GMFC-MLD“ erfolgt aus Transparenzgründen eine Darstellung im Anhang. Angesichts einer unklaren Operationalisierung und fehlender Informationen zur Validität wird auf eine Berücksichtigung des Endpunkts „Neuropsychologische Tests“ verzichtet. Zudem liegen die Rücklaufquoten des Endpunkts „Entwicklungsquotient“ bereits zu Baseline unter 70 %. Daher wird der Endpunkt in der Nutzenbewertung nicht berücksichtigt. Der Endpunkt „ARSA-Aktivität“ wird für die

Nutzenbewertung aufgrund einer fehlenden unmittelbaren Patientenrelevanz ebenfalls nicht herangezogen.

Da für den Endpunkt „GMFM“ die Rücklaufquoten nach Jahr 4 für beide Teilpopulationen unter 70 % lagen, werden nur die Ergebnisse zu Jahr 4 berichtet.

- Präsymptomatische MLD

Zu Baseline zeigte sich für den GMFM-Gesamtscore (in %) im Median ein Wert von 62 % (min; max: 13; 99). 4 Jahre nach Baseline lag der Median bei 95 % (min; max: 23; 100). Der Gesamtwert nahm also im Vergleich zu Baseline numerisch zu und näherte sich den grobmotorischen Fähigkeiten eines gesunden 5-jährigen Kindes an. Dieses Muster zeigt sich auch für die Ergebnisse der einzelnen Dimensionen A–E des GMFM. Es ist zu berücksichtigen, dass zu Jahr 4 schätzungsweise etwa 50 % der Teilnehmenden bereits mindestens 5 Jahre alt waren. Zudem liegt zu Jahr 4 ein recht hoher Anteil fehlender Werte vor.

- Frühsymptomatische MLD

Zu Baseline zeigte sich für den Gesamtwert (in %) ein Median von 87 % (min; max: 71; 99). 4 Jahre nach Baseline wurde ein Median von 46 % (min; max: 11; 64) berichtet. Die motorischen Fähigkeiten nahmen somit auf rein deskriptiver Ebene im Vergleich zu Baseline ab. Dieses Muster zeigt sich auch für die Ergebnisse der einzelnen Dimensionen A–E des GMFM. Die frühsymptomatischen Teilnehmenden waren zu Baseline im Median bereits knapp 6 Jahre alt. Zudem ist zu berücksichtigen, dass zu Jahr 4 ein recht hoher Anteil fehlender Werte vorliegt.

Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass das Instrument Funktionen abfragt, die von einem gesunden, motorisch normal entwickelten 5-jährigen Kind ausgeübt werden können.

Die Least-Square-Mittelwerte, die im Rahmen der Vorbewertung für den GMFM berichtet wurden, blieben zu Jahr 2 und 3 im Vergleich zu Baseline relativ stabil. Die für die Neubewertung vorliegende Aufteilung in eine prä- und frühsymptomatische Teilpopulation zeigt hingegen auf, dass präsymptomatische Patientinnen und Patienten im Studienverlauf im Vergleich zu Baseline numerisch bessere Werte aufwiesen, während frühsymptomatische Kinder numerisch schlechtere Werte aufwiesen. Durch die gepoolte Betrachtung beider Teilpopulationen könnten sich diese numerischen Unterschiede in der Vorbewertung gegenseitig ausgeglichen haben.

Es liegen keine Angaben zur Begleitmedikation für das IDS vor. In der pivotalen Studie 201222 erhielten insgesamt 10 % der Kinder in der Phase vor der Behandlung sowie 70 % der Kinder zu einem Zeitpunkt während des Follow-up-Zeitraums nach der Gentherapie medikamentöse Begleittherapien, die auf das Muskel-Skelett-System wirken, darunter Ibuprofen, Botulinumtoxin Typ A oder Baclofen. In den Studien HE 205029 und CUP 206258 erhielten während des Follow-up-Zeitraums nach der Gentherapie insgesamt 38 % der Kinder eine solche Begleitmedikation. Wie sich diese Kinder auf die beiden Teilpopulationen verteilen, ist nicht ersichtlich. Auswirkungen der Einnahme von Begleitmedikation, die auf das Muskel-Skelett-System wirken, auf die Ergebnisse des Endpunkts „GMFM“, sind nicht auszuschließen. Hierbei ist zu beachten, dass für die Studie 201222 nur unvollständige Angaben zur Begleitbehandlung gemacht werden können (siehe Anhang Kapitel „Begleitmedikation auf Einzelstudienebene“).

Der Effekt von OTL-200 auf die Morbidität kann auf Grundlage der vorgelegten einarmigen Studiendaten mit einhergehendem hohen Verzerrungspotential nicht abschließend beurteilt werden.

4.6 Lebensqualität

Es wurden keine Endpunkte der Kategorie „Lebensqualität“ erhoben.

4.7 Sicherheit

Für die Nutzenbewertung werden sowohl der Zeitraum a) vor der Behandlung (ab Unterzeichnung der Einwilligungserklärung bis Tag vor Beginn der Konditionierungstherapie) als auch b) ab Beginn der Konditionierung bis zum Zeitpunkt des jeweiligen Datenschnitts als relevant angesehen. Bereits vor der Konditionierung mit Busulfan erfolgte die Entnahme einer Stammzellreserve, die zwangsläufig zur Herstellung von OTL-200 notwendig ist und somit als Teil der Behandlung gewertet wird. Daher werden auch UE, die vor dem Beginn der Konditionierungstherapie, beispielsweise während der Entnahme der Stammzellreserve aufgetreten sind, als relevant für die Nutzenbewertung angesehen. Für das aktuelle Dossier reicht der pU jedoch lediglich Ergebnisse für den Zeitraum „Beginn der Konditionierung bis zum Zeitpunkt des jeweiligen Datenschnitts“ ein, welche in der Nutzenbewertung dargestellt werden. Für die Vorbewertung [10, 28], der Datenschnitte von 2018 zugrunde lagen, wurden Ergebnisse für den Zeitraum a) ab Studienbeginn bis zur Konditionierung hingegen vorgelegt. Hierbei wurde für alle 29 Patientinnen und Patienten des IDS mindestens ein UE berichtet, für 24 % mindestens 1 schweres UE und für 7 % mindestens ein SUE. Die Angaben zu den vorgenannten Phasen a) und b) beziehen sich auf das Neu- oder Wiederauftreten von UE innerhalb der jeweiligen Zeiträume. Somit werden UE, die vor der Konditionierung mit Busulfan auftraten und mit Beginn der Konditionierung nicht abgeklungen sind, nur in der ersten Studienphase erfasst. Die vorliegenden Sicherheitsergebnisse sind nur zu Phase b) und somit unvollständig.

Eine konkrete endpunktspezifische Beobachtungsdauer für die unerwünschten Ereignisse im IDS kann den eingereichten Unterlagen nicht entnommen werden. In Modul 4 wird eine endpunktunabhängige mediane Beobachtungsdauer von 10,2 Jahren (min; max: 1,14; 15,2) in der präsymptomatischen und von 10,9 Jahren (min; max: 0,64; 13,9) in der frühsymptomatischen Teilpopulation berichtet. Gemäß SAP des IDS wird die Beobachtungsdauer berechnet als der Zeitraum ab Genterapie bis zur abgeschlossenen letzten Visite. Ob dabei nur die Gabe der Infusion selbst oder auch die Konditionierung vor der Infusion berücksichtigt wird, ist nicht ersichtlich. Während für die Studien CUP 206258 und HE 205029 eine Nachbeobachtungsdauer von bis zu 8 Jahren vorgesehen war, konnten Teilnehmende der Studien 201222 und CUP 207394 auch darüber hinaus weiterbeobachtet werden. Bei beiden Studien war jedoch nach Jahr 8 lediglich eine jährliche Sicherheitserhebung vorgesehen. Es ist somit fraglich, inwiefern Sicherheitsereignisse im IDS nach Jahr 8 vollständig erfasst werden konnten.

Dem Vorgehen des pU, das Auftreten eines Todesfalls mit einem UE, das zum Therapieabbruch führte, gleichzusetzen, wird nicht gefolgt. Das Auftreten eines UE kann, je nach Zeitpunkt des Auftretens, dazu führen, dass die Konditionierung mit Busulfan nicht begonnen wird oder die OTL-200-Infusion nach der Konditionierungsbehandlung nicht verabreicht wird. Durch den pU werden derlei Therapieabbrüche aufgrund von UE bei der Darstellung allerdings nicht berücksichtigt. Gemäß Studienbericht der pivotalen Studie 201222 brachen 2 von 31 Patientinnen und Patienten die Studie vor Behandlungsbeginn ab. Somit könnten maximal 2 Personen von einem Abbruch aufgrund von UE vor der Behandlung betroffen sein.

Alle nachfolgend berichteten Ergebnisse beziehen sich auf den Zeitraum „Beginn der Konditionierung bis zum Zeitpunkt des jeweiligen Datenschnitts“. In beiden Teilpopulationen trat in diesem Zeitraum bei allen Teilnehmenden mindestens 1 UE unabhängig vom Schweregrad sowie mindestens 1 schweres UE (CTCAE-Grad ≥ 3) auf. Von den

präsymptomatischen Patientinnen und Patienten zeigten 85 % mindestens ein SUE, von den frühsymptomatischen 100 %. Somit zeigten sich zu den aktualisierten Datenschnitten numerisch mehr SUE im Vergleich zu den Datenschnitten von 2018, welche zur Vorbewertung vorlagen (69 % der IDS-Gesamtpopulation der Vorbewertung mit N = 29 hatten mindestens 1 SUE). Die Ergebnisse zu den Gesamtraten der UE und schweren UE sind identisch (auch in der Vorbewertung je 100 %).

Die häufigsten UE nach Systemorganklasse waren „Erkrankungen des Gastrointestinaltrakts“ (100 % in beiden Teilpopulationen), „Infektionen und parasitäre Erkrankungen“ (präsymptomatisch: 95 %; frühsymptomatisch: 89 %), „Untersuchungen“ (präsymptomatisch: 90 %; frühsymptomatisch: 78 %), „Erkrankungen des Blutes und des Lymphsystems“ (präsymptomatisch: 75 %; frühsymptomatisch: 89 %) sowie „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ (präsymptomatisch: 75 %; frühsymptomatisch: 100 %). Diese SOC waren unabhängig vom UE-Schweregrad auch unter den häufigsten berichteten SOC zu den Datenschnitten von 2018 in der Vorbewertung.

Die häufigsten Systemorganklassen bei den schweren UE waren „Erkrankungen des Gastrointestinaltrakts“ (präsymptomatisch: 80 %; frühsymptomatisch: 100 %), „Erkrankungen des Blutes und des Lymphsystems“ (präsymptomatisch: 75 %; frühsymptomatisch: 89 %), „Infektionen und parasitäre Erkrankungen“ (präsymptomatisch: 65 %; frühsymptomatisch: 44 %), „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ (präsymptomatisch: 55 %; frühsymptomatisch: 78 %) sowie „Erkrankungen des Nervensystems“ (präsymptomatisch: 55 %; frühsymptomatisch: 100 %). Auch diese Ergebnisse sind weitestgehend analog zu den Ergebnissen der Datenschnitte von 2018.

Die häufigsten Systemorganklassen bei den SUE waren „Erkrankungen des Gastrointestinaltrakts“ (präsymptomatisch: 40 %; frühsymptomatisch: 33 %), „Infektionen und parasitäre Erkrankungen“ (präsymptomatisch: 40 %; frühsymptomatisch: 22 %), „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ (präsymptomatisch: 25 %; frühsymptomatisch: 56 %) sowie „Erkrankungen des Nervensystems“ (präsymptomatisch: 25 %; frühsymptomatisch: 67 %). Die SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ war zu den Datenschnitten von 2018 noch nicht unter den häufigsten SOC bei den SUE.

Aus der deskriptiven Darstellung in Tabelle 19 und Tabelle 20 lässt sich ableiten, dass beispielsweise alle UE, die mit den PT „Motorische Funktionsstörung“ (präsymptomatisch: 25 %; frühsymptomatisch: 89 %), „Aphasie“ (präsymptomatisch: 20 %; frühsymptomatisch: 67 %), „Ataxie“ (präsymptomatisch: 10 %; frühsymptomatisch: 56 %), „Kognitive Störung“ (präsymptomatisch: 35 %; frühsymptomatisch: 67 %) (SOC „Erkrankungen des Nervensystems“) kodiert wurden oder aber mit dem PT „Gangstörung“ (präsymptomatisch: 50 %; frühsymptomatisch: 78 %) (SOC „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“), UE mit CTCAE-Grad ≥ 3 waren. Die letztgenannten UE auf Ebene der PT spiegeln krankheitsbedingte Symptome wider, die seitens des pU ab einem CTCAE-Grad 3 als UE gewertet wurden. Der pU legte keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung mit in die Erfassung der UE eingingen. Insbesondere für die SOC „Erkrankungen des Nervensystems“ kann nicht ausgeschlossen werden, dass Symptome der Endpunktkategorie „Morbidität“ subsumiert wurden. Eine Präspezifikation, welche Symptome bzw. SOC und PT konkret der Grunderkrankung zugeordnet wurden, kann den Studienunterlagen nicht entnommen werden. Entsprechend kann eine potentielle Untererfassung der UE mit Schweregrad < 3 nicht ausgeschlossen werden.

Gemäß dem 2020 zur Erstzulassung von OTL-200 veröffentlichten EPAR [4] sollte die Sicherheitsnachbeobachtung bei Gentherapien 15 Jahre betragen. Da dies durch die

Nachbeobachtungszeiträume der IDS-Studien nicht ausreichend abgedeckt wird, ordnete die EMA eine Langzeitnachbeobachtungsstudie (LongTERM-MLD) an. In diese Studie sollen sowohl Patientinnen und Patienten aus dem klinischen Entwicklungsprogramm als auch neue Patientinnen und Patienten aufgenommen werden, die gemäß einem vereinbarten Protokoll mit der kommerziellen Formulierung behandelt werden. Zum Zeitpunkt der vorliegenden Nutzenbewertung liegen jedoch noch keine Ergebnisse der LongTERM-MLD vor (siehe auch Tabelle 1).

Der Effekt von OTL-200 auf die Sicherheit kann auf Grundlage der vorgelegten einarmigen Studiendaten mit einhergehendem hohen Verzerrungspotential nicht abschließend beurteilt werden.

5 Anforderung an eine qualitätsgesicherte Anwendung

Die Angaben der Fachinformation sind zu berücksichtigen.

6 Zusammenfassung der Nutzenbewertung

Atidarsagen autotemcel/OTL-200 (Libmeldy®), im Folgenden als OTL-200 bezeichnet, ist zugelassen für Metachromatische Leukodystrophie (MLD), die durch Mutationen in beiden Allelen des Gens für die Arylsulfatase A (ARSA) gekennzeichnet ist, welche zu einer Verringerung der enzymatischen Aktivität von ARSA führt:

- a) bei Kindern mit im späten Säuglings- oder frühen Kindesalter auftretenden Formen, ohne klinische Manifestation der Erkrankung,
- b) bei Kindern mit der im frühen Kindesalter auftretenden Form mit frühzeitiger klinischer Manifestation der Erkrankung, die jedoch noch selbstständig gehen können, vor dem Einsetzen einer kognitiven Verschlechterung.

Die Nutzenbewertung von OTL-200 basiert auf der zulassungsbegründenden Studie 201222 sowie den Compassionate-Use- und Hospital-Exemption-Programmen CUP 207394, CUP 206258 und HE 205029. Die 4 Studien wurden gemeinsam in Form eines integrierten Datensatzes (IDS) ausgewertet. Es handelt sich bei der Studie 201222 um eine nicht-randomisierte, einarmige Studie, bei der Studie CUP 207394 um einen Fallbericht und bei den Studien CUP 206258 und HE 205029 jeweils um eine Fallserie.

Weitergehende, über die Angaben in der nachfolgenden zusammenfassenden Tabelle hinausgehende Erläuterungen zur Methodik und den Ergebnissen der Endpunkte finden sich in den jeweiligen Kapiteln 2.3 bis 2.5 und 3.2 bis 3.5 der Nutzenbewertung.

Tabelle 22: Zusammenfassende Darstellung der Ergebnisse des IDS

IDS Darstellung der Ergebnisse	OTL-200 Präsymptomatisch N = 20		OTL-200 Frühsymptomatisch N = 9	
Mortalität				
Gesamtüberleben	N ¹⁾	Personen mit Ereignis, n (%)	N ¹⁾	Personen mit Ereignis, n (%)
Tod	20	1 (5)	9	2 (22)
Morbidität				
GMFM-Score ²⁾ (%)	N ¹⁾	Median (min; max)	N ¹⁾	Median (min; max)
Baseline	20	61,6 (13,4; 98,6)	9	86,6 (71,1; 98,7)
Jahr 4 ³⁾	15	95 (22,7; 100)	5	45,8 (10,9; 64)
Sicherheit				
Unerwünschte Ereignisse ⁴⁾	N ¹⁾	Personen mit Ereignis, n (%)	N ¹⁾	Personen mit Ereignis, n (%)
Schwere UE ⁵⁾ Studienbeginn bis Konditionierung ⁶⁾ Beginn Konditionierung bis Datenschnitt ⁷⁾	20	k. A. 20 (100)	9	k. A. 9 (100)
SUE Studienbeginn bis Konditionierung ⁶⁾ Beginn Konditionierung bis Datenschnitt ⁷⁾	20	k. A. 17 (85)	9	k. A. 9 (100)

IDS Darstellung der Ergebnisse	OTL-200 Präsymptomatisch N = 20		OTL-200 Frühsymptomatisch N = 9	
UE, das zum Abbruch der Studienmedikation führte ⁸⁾	20		9	
Studienbeginn bis Konditionierung ⁶⁾		k. A.		k. A.
Beginn Konditionierung bis Datenschnitt ⁷⁾		k. A.		k. A.

¹⁾ Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden.

²⁾ Der Wertebereich der GMFM reicht von 0–100 %, wobei ein höherer Wert besser ausgeprägten grobmotorischen Funktionen entspricht.

³⁾ Zu Jahr 4 waren 3 Personen, davon 1 mit präsymptomatischer MLD und 2 mit frühsymptomatischer MLD, bereits verstorben.

⁴⁾ Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind in Kapitel 3.5 dargestellt.

⁵⁾ Schweregrad mittels der Common Terminology Criteria of Adverse Event (CTCAE) Version 3.0 klassifiziert.

⁶⁾ Zeitraum ab Tag der Unterzeichnung der Einwilligungserklärung bis zum Tag vor Beginn der Konditionierung mit Busulfan. UE, die ab Unterzeichnung der Einwilligungserklärung bis zum Tag vor Beginn der Konditionierungstherapie auftraten und ggf. weiterbestanden haben, sind nicht in den Ergebnissen berücksichtigt.

⁷⁾ Datenschnitte vom 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029).

⁸⁾ Dem Vorgehen des pU, das Auftreten eines Todesfalls mit einem UE, das zum Therapieabbruch führte, gleichzusetzen, wird nicht gefolgt. Das Auftreten von UE kann zur Folge haben, dass die Konditionierung nicht begonnen wird oder die OTL-200-Infusion nach der Konditionierung nicht verabreicht wird. Eine Darstellung ebendieser Therapieabbrüche wegen UE erfolgt seitens des pU nicht. Es wird davon ausgegangen, dass dies maximal 2 Personen betreffen konnte.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; GMFM: Gross Motor Function Measure; IDS: Integrierter Datensatz; k. A.: keine Angabe; MLD: Metachromatische Leukodystrophie; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Referenzen

1. **American Psychiatric Association (APA).** Diagnostic and statistical manual of mental disorders, 4th edition (DSM IV). Washington DC: APA; 1994.
2. **American Psychiatric Association (APA).** Diagnostische Kriterien DSM-5-TR. Göttingen (GER): Hogrefe Verlag; 2025.
3. **European Medicines Agency (EMA).** Clinical-response to EMA day 120 list of questions EMEA/H/C/005321, Libmeldy (OTL-200) [unveröffentlicht].
4. **European Medicines Agency (EMA).** Libmeldy: European public assessment report EMEA/H/C/005321/0000 [online]. 15.10.2020. Amsterdam (NED): EMA. [Zugriff: 01.07.2026]. URL: https://www.ema.europa.eu/en/documents/assessment-report/libmeldy-epar-public-assessment-report_en.pdf.
5. **European Medicines Agency (EMA).** Libmeldy: Zusammenfassung der Merkmale des Arzneimittels (EPAR Product information) [online]. Last update: 16.12.2025. Amsterdam (NED): EMA; 2020. [Zugriff: 02.07.2026]. URL: https://www.ema.europa.eu/de/documents/product-information/libmeldy-epar-product-information_de.pdf.
6. **Gemeinsamer Bundesausschuss (GBA).** Amendment 1 zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO Wirkstoff: Atidarsagen autotemcel/OTL-200 [online]. Berlin (GER): GBA; 2021. [Zugriff: 09.09.2026]. URL: https://www.g-ba.de/downloads/92-975-5005/2021-11-04_Amendment-G-BA_Atidarsagen-autotemcel_D-678.pdf.
7. **Gemeinsamer Bundesausschuss (GBA).** Amendment 2 zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO Wirkstoff: Atidarsagen autotemcel/OTL-200 [online]. Berlin (GER): GBA; 2021. [Zugriff: 09.09.2026]. URL: https://www.g-ba.de/downloads/92-975-5004/2021-11-04_Amendment-2-G-BA_Atidarsagen-autotemcel_D-678.pdf.
8. **Gemeinsamer Bundesausschuss (GBA).** Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Atidarsagen autotemcel (Metachromatische Leukodystrophie mit biallelischer Mutation im ARSA-Gen) vom 4. November 2021 [online]. Berlin (GER): GBA; 2021. [Zugriff: 18.09.2026]. URL: https://www.g-ba.de/downloads/39-261-5103/2021-11-04_AM-RL-XII_Atidarsagen-autotemcel_D-678.pdf.
9. **Gemeinsamer Bundesausschuss (GBA).** Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Änderung der Angaben zur Geltungsdauer eines Beschlusses über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V) Atidarsagen autotemcel (Metachromatische Leukodystrophie mit biallelischer Mutation im ARSA-Gen) vom 20. Juni 2024 [online]. Berlin (GER): GBA; 2024. [Zugriff: 18.09.2026]. URL: https://www.g-ba.de/downloads/39-261-6662/2024-06-20_AM-RL-XII_Atidarsagen-autotemcel_D-678_BAnz.pdf.

10. **Gemeinsamer Bundesausschuss (GBA).** Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO Wirkstoff: Atidarsagen autotemcel/OTL-200 [online]. Berlin (GER): GBA; 2021. [Zugriff: 09.09.2026]. URL: https://www.g-ba.de/downloads/92-975-4677/2021-08-01_Nutzenbewertung_G-BA_D-678.pdf.
11. **Gemeinsamer Bundesausschuss (GBA).** Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Atidarsagen autotemcel (Metachromatische Leukodystrophie mit biallelischer Mutation im ARSA-Gen) vom 4. November 2021 [online]. Berlin (GER): GBA; 2021. [Zugriff: 18.09.2026]. URL: https://www.g-ba.de/downloads/40-268-7994/2021-11-04_AM-RL-XII_Atidarsagen-autotemcel_D-678_TrG.pdf.
12. **Gemeinsamer Bundesausschuss (GBA).** Zusammenfassende Dokumentation über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Atidarsagen autotemcel OTL-200 vom 04. November 2021 [online]. Berlin (GER): GBA; 2021. [Zugriff: 15.09.2026]. URL: https://www.g-ba.de/downloads/40-268-8314/2021-11-04_AM-RL-XII_Atidarsagen-autotemcel_D-678_ZD.pdf.
13. **GlaxoSmithKline.** Compassionate use program for hematopoietic stem cell gene therapy (GSK2696274) in pre-symptomatic late infantile metachromatic leukodystrophy patients; clinical protocol [unveröffentlicht]. 06.01.2017.
14. **GlaxoSmithKline.** Compassionate use program for hematopoietic stem cell gene therapy (GSK2696274) in pre-symptomatic late infantile metachromatic leukodystrophy patients; protocol number: 206258 amendment 01 [unveröffentlicht]. 12.07.2017.
15. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).** Neugeborenencreening auf metachromatische Leukodystrophie; Abschlussbericht S25-01 [online]. Köln (GER): IQWiG; 2026. [Zugriff: 09.09.2026]. (IQWiG-Berichte; 2273). URL: https://www.iqwig.de/download/s25-01_neugeborenencreening-auf-metachromatische-leukodystrophie_abschlussbericht_v1-0.pdf.
16. **Kehrer C, Blumenstock G, Raabe C, Krägeloh-Mann I.** Development and reliability of a classification system for gross motor function in children with metachromatic leucodystrophy. Dev Med Child Neurol 2011;53(2):156-160. <https://dx.doi.org/10.1111/j.1469-8749.2010.03821.x>.
17. **MolMed.** Hematopoietic stem cell gene therapy for presymptomatic late infantile metachromatic leukodystrophy prepared on non-routine basis for patient coded MLD-HE01; protocol, vers. 1 [unveröffentlicht]. 19.11.2015.
18. **MolMed.** Hematopoietic stem cell gene therapy for presymptomatic late infantile metachromatic leukodystrophy prepared on non-routine basis for patient coded MLD-HE02; protocol, vers. 1 [unveröffentlicht]. 19.11.2015.
19. **MolMed.** Hematopoietic stem cell gene therapy for presymptomatic late infantile metachromatic leukodystrophy prepared on non-routine basis for patient coded MLD-HE02; protocol, vers. 2 [unveröffentlicht]. 31.03.2016.
20. **MolMed.** Hematopoietic stem cell gene therapy for presymptomatic late infantile metachromatic leukodystrophy prepared on non-routine basis for patient coded MLD-HE03; protocol, vers. 1 [unveröffentlicht]. 19.11.2015.

21. **Orchard Therapeutics.** Clinical program report (205029/206258) for Expanded Access Program (EAP) for hematopoietic stem cell gene therapy OTL-200 in subjects with early-onset metachromatic leukodystrophy (MLD) - Hospital Exemption (HE) protocols 205029 and Compassionate Use Program (CUP) 206258, tables and figures [unveröffentlicht]. 2025.
22. **Orchard Therapeutics.** Clinical program report (205029/206258) for Expanded Access Program (EAP) for hematopoietic stem cell gene therapy OTL-200 in subjects with early-onset metachromatic leukodystrophy (MLD) - Hospital Exemption (HE) protocols 205029 and Compassionate Use Program (CUP) 206258, vers. 3.0 [unveröffentlicht]. 19.12.2025.
23. **Orchard Therapeutics.** Compassionate Use Program for haematopoeitic stem cell gene therapy OTL-200 (formerly known as GSK2696274) in pre-symptomatic early onset metachromatic leukodystrophy patients, protocol number: 206258 amendment 02 [unveröffentlicht]. 31.07.2018.
24. **Orchard Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 1: Atidarsagen autotemcel/OTL-200 (Libmeldy), Zusammenfassung der Aussagen im Dossier [unveröffentlicht]. 30.06.2026.
25. **Orchard Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 2: Atidarsagen autotemcel/OTL-200 (Libmeldy), Allgemeine Angaben zum Arzneimittel, zugelassene Anwendungsgebiete [unveröffentlicht]. 30.06.2026.
26. **Orchard Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 3 A: Atidarsagen autotemcel/OTL-200 (Libmeldy), Behandlung von früheinsetzender metachromatischer Leukodystrophie (MLD); Zweckmäßige Vergleichstherapie, Anzahl der Patienten mit therapeutisch bedeutsamem Zusatznutzen, Kosten der Therapie für die GKV, Anforderungen an eine qualitätsgesicherte Anwendung, Prüfungsteilnehmer im Geltungsbereich des SGB V [unveröffentlicht]. 30.06.2026.
27. **Orchard Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Atidarsagen autotemcel (Libmeldy), Behandlung von früheinsetzender metachromatischer Leukodystrophie (MLD); Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 30.06.2026.
28. **Orchard Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Atidarsagen autotemcel/OTL-200 (Libmeldy), Behandlung der metachromatischen Leukodystrophie (MLD) bei Kindern; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [online]. Berlin (GER): Gemeinsamer Bundesausschuss (GBA); 01.05.2021. [Zugriff: 15.09.2026]. URL: https://www.g-ba.de/downloads/92-975-4674/2021_05_01_Modul4A_Atidarsagen_autotemcel_OTL200.pdf.
29. **Orchard Therapeutics.** Gene therapy protocol using autologous hematopoietic stem cells for MLD-C02, a patient with metachromatic leukodystrophy (MLD) - Compassionate Use Program 207394; clinical report [unveröffentlicht]. 2026.
30. **Orchard Therapeutics.** Gene therapy protocol using autologous hematopoietic stem cells for MLD-C02, a patient with metachromatic leukodystrophy (MLD) (study 207394); clinical program report, figures and tables [unveröffentlicht].
31. **Orchard Therapeutics.** Gene therapy protocol using autologous hematopoietic stem cells for MLD-C02, a patient with metachromatic leukodystrophy (MLD) (study 207394); clinical program report, vers. 4.0 [unveröffentlicht]. 27.01.2026.

32. **Orchard Therapeutics.** Gene therapy protocol using autologous hematopoietic stem cells for MLD-C02, a patient with metachromatic leukodystrophy (MLD) (study 207394); protocol, vers. 3.0 [unveröffentlicht]. 12.02.2018.
33. **Orchard Therapeutics.** Gene therapy protocol using autologous hematopoietic stem cells for MLD-C02, a patient with metachromatic leukodystrophy (MLD) (study 207394); statistical analysis plan, vers. 5.0 [unveröffentlicht]. 26.08.2026.
34. **Orchard Therapeutics.** Long-term efficacy and safety follow-up of metachromatic leukodystrophy patients treated with Atidarsagene Autotemcel (study OTL_200_10); clinical study report, vers. 4.3 [unveröffentlicht]. 10.02.2025.
35. **Orchard Therapeutics.** OTL-200; integrated statistical analysis plan [unveröffentlicht]. 17.09.2019.
36. **Orchard Therapeutics.** A phase I/II clinical trial of hematopoietic stem cell gene therapy for the treatment of metachromatic leukodystrophy (study 201222); clinical study report [unveröffentlicht]. 2026.
37. **Orchard Therapeutics.** A phase I/II clinical trial of hematopoietic stem cell gene therapy for the treatment of metachromatic leukodystrophy (study 201222); clinical study report, tables and figures [unveröffentlicht]. 2025.
38. **Orchard Therapeutics.** A phase I/II clinical trial of hematopoietic stem cell gene therapy for the treatment of metachromatic leukodystrophy (study 201222); protocol v13.3 [unveröffentlicht]. 28.02.2025.
39. **Orchard Therapeutics.** A phase I/II clinical trial of hematopoietic stem cell gene therapy for the treatment of metachromatic leukodystrophy (study 201222); statistical analysis plan v6 [unveröffentlicht]. 26.08.2025.
40. **Orchard Therapeutics.** A single arm, open label clinical study of cryopreserved autologous CD34+ cells transduced with lentiviral vector containing human ARSA cDNA (OTL-200) for the treatment of early onset metachromatic leukodystrophy (MLD) (study 205756); clinical study report [unveröffentlicht]. 26.05.2023.
41. **Orchard Therapeutics.** Statistical analysis plan for meta-analysis of expanded access programs 205029/206258, vers. 3.0 [unveröffentlicht]. 2025.
42. **Russell D, Rosenbaum P, Avery LM, Lane M.** Gross Motor Function Measure (GMFM-66 and GMFM-88) user's manual. London: Mac Keith Press; 2002.

Anhang

Begleitmedikation auf Einzelstudienebene

Für die Studie CUP 207394 finden sich in den Studienunterlagen keine Daten zu (nicht-) medikamentösen Begleittherapien. In der pivotalen Studie 201222 sowie den Studien HE 205029 und CUP 206258 werden zwar keine Angaben zu den Begleittherapien für den gesamten Studienzeitraum ab Unterzeichnung der Einverständniserklärung bis Studienende gemacht. Stattdessen werden jedoch die Begleittherapien für unterschiedliche Studienabschnitte separat berichtet: Vor der Behandlung, während der Behandlung sowie für den gesamten Follow-up-Zeitraum nach der Behandlung. Die Zeiträume „vor der Behandlung“ und „nach der Behandlung“ sind dabei nicht näher definiert; es ist unklar, ob „vor der Behandlung“ den Zeitraum der Busulfan-Konditionierung umfasst oder ob die Konditionierung dem Zeitraum „während der Behandlung“ zugeordnet wird. Zudem kann für beide Zeiträume kein konkreter Startzeitpunkt identifiziert werden.

In der pivotalen Studie 201222 (N = 20) erhielten insgesamt 10 % der Kinder in der Phase vor der Behandlung sowie 70 % der Kinder zu einem Zeitpunkt während des Follow-up-Zeitraums nach der Behandlung medikamentöse Begleittherapien, die auf das Muskel-Skelett-System wirken, darunter Ibuprofen (vor der Behandlung 10 % sowie während des Follow-up-Zeitraums 40 %), Botulinumtoxin Typ A (während des Follow-up-Zeitraums, 40 %) oder Baclofen (während des Follow-up-Zeitraums, 40 %). In den Studien HE 205029 und CUP 206258 (N = 8) erhielten während des Follow-up-Zeitraums nach der Behandlung insgesamt 38 % der Kinder eine solche Begleitmedikation, darunter ebenfalls Ibuprofen (vor der Behandlung, 13 %; während des Follow-up-Zeitraums, 25 %), Botulinumtoxin Typ A (während des Follow-up-Zeitraums, 13 %) und Baclofen (während des Follow-up-Zeitraums, 13 %).

Im Studienprotokoll der Studie 201222 wird beschrieben, dass einige Medikamente, wie beispielsweise Antiemetika, Analgetika und Antiepileptika, routinemäßig im Studienzentrum eingesetzt werden. Sofern diese wegen einer Begleiterkrankung oder UE zum Einsatz kamen, wurde dies erfasst. Die routinemäßige Anwendung wurde hingegen nicht dokumentiert. Somit ist zu beachten, dass für die Studie 201222 nur unvollständige Angaben zur Begleitbehandlung gemacht werden können.

GMFC-MLD

Tabelle 23: Ergebnisse für „Anteil der Personen mit einem GMFC-MLD-Level ≥ 5 zum Zeitpunkt des letzten Follow-ups“ im IDS zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258) und 26.05.2025 (HE 205029); ITT-Population

IDS GMFC-MLD ¹⁾	OTL-200 Präsymptomatisch N = 20	OTL-200 Frühsymptomatisch N = 9
Anteil der Personen mit einem GMFC-MLD-Level ≥ 5 zum letzten Follow-up²⁾		
GMFC-MLD-Level ≥ 5 , n (%) ³⁾		
Ja	2 (10)	5 (56)
Nein ⁴⁾	18 (90)	4 (44)
Beobachtungsdauer ⁵⁾ (in Jahren), Median (min; max)	10,2 (1,14; 15,2)	10,9 (0,64; 13,9)

¹⁾ Klassifikationssystem zur Beschreibung grobmotorischer Funktionen von Kindern mit MLD, die mindestens 18 Monate alt sind. Das Klassifikationssystem umfasst 7 verschiedene Stufen. GMC-MLD-Level ≥ 5 : Verlust der Fähigkeiten, sich selbstständig fortbewegen und selbstständig sitzen zu können; Kopfkontrolle erhalten.

²⁾ Post-hoc-Analyse.

³⁾ Gemäß Modul 4 soll zwar bei fehlenden Werten eine „Worst-Case-Imputation“ durchgeführt werden, bei welcher ein fehlender Wert als Non-Response gewertet wird. Dabei ist zum einen unklar, ob dies für alle Arten von fehlenden Werten (inkl. aufgrund von Todesfällen) galt. Zum anderen kann dies jedoch insbesondere inhaltlich nicht nachvollzogen werden, da die Angaben im Zuge der Auswertungen implizieren, dass eine Response ab einem Level 5 vorlag und höhere Werte beim GMFC-MLD eine größere Einschränkung der grobmotorischen Fähigkeiten bedeuten. Demnach entspräche die Wertung von fehlenden Werten als Response einer „Worst-Case-Imputation“ und nicht – wie in Modul 4 beschrieben – die Wertung von fehlenden Werten als Non-Response.

⁴⁾ Es ist unklar, für wie viele Patientinnen und Patienten zum (nicht näher definierten) Zeitpunkt des letzten Follow-ups fehlende Werte vorlagen. Da der Umgang mit fehlenden Werten nicht nachvollzogen werden kann, kann nicht ausgeschlossen werden, dass fehlende Werte als „Ja“ oder „Nein“ kategorisiert wurden, obwohl das GMFC-MLD-Level zum letzten Follow-up unbekannt war.

⁵⁾ Gemäß SAP ist die Beobachtungsdauer definiert als die Zeit von der Gentherapie bis zum Ende der letzten Studienvisite.

Abkürzungen: GMFC-MLD: Gross Motor Function Classification in MLD; IDS: Integrierter Datensatz; ITT: Intention-to-Treat; MLD: Metachromatische Leukodystrophie; SAP: statistischer Analyseplan.

Geschwisterkinderanalyse IDS vs. TIGET-NHx

Für die Studien des IDS sind in Modul 4 folgende Datenschnitte angegeben:

- 201222: 30.06.2025
- CUP 207394: 30.06.2025
- CUP 206258: 26.05.2025
- HE 205029: 26.05.2025

Für die TIGET-NHx-Studie ist zudem im Studienbericht ein Datenschnitt vom 1. November 2021 angegeben.

Die Rekrutierung der pivotalen Studie wurde 2016 bereits abgeschlossen. Die ITT-Population des IDS beträgt, analog zur ITT-Population, die für die Vorbewertung vorgelegt wurde, weiterhin N = 29 Patientinnen und Patienten. Die Geschwisterkinderanalyse (IDS vs. TIGET-NHx), die der pU für die Vorbewertung vorgelegt hatte, umfasste jedoch N = 11 mit OTL-200 behandelte Kinder und N = 10 unbehandelte Kinder der Studie TIGET-NHx, während die aktuell vorliegende Geschwisterkinderanalyse N = 17 behandelte und N = 15 unbehandelte Kinder umfasst (siehe Tabelle 24). Warum nun mehr Kinder für den Vergleich zur Verfügung standen, ist unklar.

Studiencharakteristika und Studienmedikation

Tabelle 24: Allgemeine Angaben; Geschwisterkinderanalyse IDS vs. TIGET-NHx, MSAS

MSAS Allgemeine Angaben	IDS (OTL-200) N = 17	TIGET-NHx (BSC) N = 15
MSAS-Population ¹⁾ , n		
Studie 201222	11	0
TIGET-NHx	0	15
Studie CUP 206258	3	0
Studie HE 205029	3	0
Studienabbruch, n (%)	k. A.	k. A.
Beobachtungsdauer (Jahre)		
n (%)	17 (100)	15 (100)
Mittelwert (SD)	10,7 (2,56)	5,36 (3,86)
Median (min; max)	10,6 (7,64; 15,2)	4,54 (1,96; 17,9)

¹⁾ Alle Patientinnen und Patienten der ITT-Population mit einem Geschwisterkind in der Studie TIGET-NHx und alle Geschwisterkinder der Studie TIGET-NHx.

Abkürzungen: BSC: Best Supportive Care; IDS: integrierter Datensatz; ITT: Intention-to-Treat; k. A.: keine Angabe; MSAS: Matched Sibling Analysis Set; SD: Standardabweichung.

Tabelle 25: Charakterisierung der Studienpopulation; Geschwisterkinderanalyse IDS vs. TIGET-NHx, MSAS

MSAS Charakterisierung der Studienpopulation	IDS (OTL-200) N = 17	TIGET-NHx (BSC) N = 15
Alter zu Krankheitsbeginn ¹⁾ (Monate)		
MW (SD)	28,7 (19,1)	27,9 (20,9)
Median (min; max)	18 (15; 75)	18 (15; 75)
Alter zu Diagnosestellung (Monate)		
MW (SD)	21,4 (32,4)	42,8 (19,6)
Median (min; max)	11 (1,88; 131,9)	35 (21,2; 85,2)

MSAS Charakterisierung der Studienpopulation	IDS (OTL-200) N = 17	TIGET-NHx (BSC) N = 15
<i>Alter zu Screening/Studieneinschluss (Monate)</i> Mittelwert (SD) Median (min; max)	24,1 (32,5) 12,9 (6,57; 136,2)	71,8 (37,2) 60,7 (34,9; 176,4)
<i>Alter zum Zeitpunkt des 1. Kontaktes²⁾ (Monate)</i> Mittelwert (SD) Median (min; max)	23,8 (32,7) 13 (3,59; 136,4)	30,7 (20,1) 20,1 (17,7; 74,1)
<i>Alter zum Zeitpunkt der Gentherapie (Monate)</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	11 (64,7) ⁴⁾ 34,1 (39,3) 17,7 (7,56; 139,7)	n. a.
<i>Alter bei letzter Untersuchung (Monate)</i> Mittelwert (SD) Median (min; max)	k. A.	k. A.
<i>Alter zu Matchingbaseline</i>	k. A.	k. A.
<i>Geschlecht, n (%)</i> Weiblich Männlich	7 (41) 10 (59)	9 (60) 6 (40)
<i>Abstammung (genetisch), n (%)</i> Weiß - kaukasisch/europäisch Weiß - arabisch/nordafrikanische Herkunft	13 (76) 4 (24)	12 (80) 3 (20)
<i>Ethnie, n (%)</i> Nicht Hispanisch oder Latino Hispanisch oder Latino	16 (94) 1 (6)	14 (93) 1 (7)
<i>MLD-Form, n (%)</i> Spätingfantil (LI) Frühjuvenil (EJ) ⁵⁾	13 (76) 4 (24)	11 (73) ⁶⁾ 4 (27)
<i>Symptomstatus zu Screening/Studieneinschluss, n (%)</i> Präsymptomatisch Frühsymptomatisch	16 (94) 1 (6)	0 (0) 15 (100)
<i>Symptomstatus zum Zeitpunkt der Gentherapie, n (%)</i> Präsymptomatisch Frühsymptomatisch	15 (88) 2 (12)	n. a.
<i>Symptomstatus zu Matchingbaseline, n (%)</i>	k. A.	k. A.
<i>GMFM-Score (%)⁷⁾ zu Baseline⁸⁾</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	17 (100) ⁴⁾ 61,3 (26,2) 65 (13,4; 98,6)	11 (73,3) ⁴⁾ 21,2 (30,4) 9,33 (1,18; 96,4)
<i>GMFC-MLD⁹⁾ zu Baseline⁸⁾, n (%)</i> Level 0 Level 1 k. A. ¹⁰⁾	3 (18) 2 (12) 12 (71)	0 4 (27) 11 (73)
<i>ARSA-Aktivität_{PBMC} (nmol/mg/h)¹¹⁾ zu Baseline⁸⁾</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	17 (100) ⁴⁾ 25,8 (0) 25,8 (25,8; 25,8)	1 (6,67) ⁴⁾ 25,8 (-) 25,8 (25,8; 25,8)

MSAS Charakterisierung der Studienpopulation	IDS (OTL-200) N = 17	TIGET-NHx (BSC) N = 15
<i>MRT-Score¹²⁾ zu Baseline⁸⁾</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	17 (100) ⁴⁾ 0,88 (1,71) 0 (0; 4,75)	11 (73,3) ⁴⁾ 16,4 (6,51) 13,5 (10,3; 28,5)
<i>NCV-Index¹³⁾ zu Baseline⁸⁾</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	17 (100) ⁴⁾ -3,94 (3,42) -3,14 (-10,3; 0,55)	7 (46,7) ⁴⁾ -13,2 (1,00) -13,6 (-14,3; -11,7)
<i>Begleitende Therapie zu Studienbeginn, n (%)</i> Ja Nein	k. A.	k. A.
<i>Vorthherapie zu Studienbeginn, n (%)</i> Ja Nein	k. A.	k. A.

¹⁾ Bei präsymptomatischen Patientinnen und Patienten (betrifft ausschließlich Kinder im OTL-200-Arm) wurde das vorhergesagte Alter zu Krankheitsbeginn auf der Grundlage des Alters beim Auftreten der ersten Symptome bei den älteren Geschwistern der Patientin / des Patienten berechnet. In Bezug auf die ITT-Population war bei 5 Kindern in der Studie 201222, bei 2 Kindern der Studie HE 205029 sowie bei einem Kind der Studie CUP 206258 das genaue Alter beim Auftreten der ersten Symptome beim unbehandelten Geschwisterkind ungewiss und wurde im eCRF als Spanne erfasst; das höhere Alter wurde als voraussichtliches Alter für den Beginn der Symptome beim behandelten Kind herangezogen. Wie viele dieser Kinder im MSAS enthalten sind, ist unklar.

²⁾ Für Kinder des OTL-200-Arms entspricht dies dem Alter zum Zeitpunkt der Baseline-Erhebung von GMFC-MLD, GMFM, NCV und MRT. Es ist unklar, um welche Baseline es sich für die Teilnehmenden der TIGET-NHx-Studie handelt.

³⁾ Eine Angabe von n (%) erfolgt nur bei Charakteristika, für die keine vollständigen Daten vorlagen.

⁴⁾ Prozentzahlen aus Eigenberechnung.

⁵⁾ Für die Analysen wurde eine Person der Studie 201222 mit einer intermediären Form der frühjuvenilen Form (EJ) zugeordnet. Ob diese Person prä- oder frühsymptomatisch war und im MSAS enthalten ist, ist unklar.

⁶⁾ Da alle Personen im TIGET-NHx-Arm frühsymptomatisch waren, entspricht diese Population nicht dem zugelassenen Anwendungsgebiet von OTL-200.

⁷⁾ Der Wertebereich der GMFM reicht von 0–100 %, wobei ein höherer Wert stärker ausgeprägten grobmotorischen Funktionen entspricht.

⁸⁾ Angabe „zu Baseline“ gemäß Modul 4. Es ist unklar, ob es sich um die Matchingbaseline handelt oder um die jeweilige Baseline-Erhebung in den IDS-Studien bzw. der TIGET-NHx-Studie. Insbesondere für die TIGET-NHx-Studie liegt zudem keine Matchingbaseline-Definition vor.

⁹⁾ Klassifikationssystem zur Beschreibung grobmotorischer Funktionen von Kindern mit MLD, die mindestens 18 Monate alt sind. Das Klassifikationssystem umfasst 7 verschiedene Stufen (Stufe 0: altersentsprechendes, freies Gehen ohne Unterstützung; Stufe 6: Verlust jeglicher Fortbewegungsfähigkeit sowie jeglicher Kopf- und Rumpfkontrolle).

¹⁰⁾ Da das GMFC-MLD-Level auf der Fähigkeit zu Gehen basiert, kann es in der Regel erst ab einem Alter von mindestens 18 Monaten bestimmt werden. Somit konnten für Kinder mit einem Alter von < 18 Monate keine Baseline-Werte erhoben werden. Bei Kindern, die jünger sind als 18 Monate, aber schon laufen können, kann das GMFC-MLD-Klassifikationssystem gemäß pU jedoch angewendet werden. Ob alle fehlenden Werte darauf zurückzuführen sind, dass die entsprechenden Kinder < 18 Monate alt waren bzw. noch nicht laufen konnten, kann den Studienunterlagen nicht entnommen werden.

¹¹⁾ Da sich die Werte für die ARSA-Aktivität unterhalb des messbaren Bereichs bewegten, wurden diese laut Modul 4 mit dem LLOQ imputiert.

¹²⁾ Der MRT-Score wurde anhand einer adaptierten Loes' Skala ermittelt, die Subskalen zu Demyelinisierung (0–16 Punkte), b) globaler (0–4 Punkte) und fokaler (0–10,5 Punkte) Hirnatrophie sowie getigertem Muster ('tigroid pattern') (0–1 Punkte) umfasst. Der Gesamtscore kann Werte von 0 (unauffällig) bis 31,5 (merklich verändert) annehmen. Ein Wert von > 0 gilt als abnormal.

¹³⁾ Der NCV-Index berechnet sich aus dem Durchschnitt von 4 z-Scores, die auf Testungen der Nervenleitgeschwindigkeit des Nervus suralis, Nervus medianus, Nervus fibularis profundus und Nervus ulnaris beruhen. Für die Berechnung der z-Scores wurden, je nach Alter der Patientinnen und Patienten, 3 altersbezogene Normkollektive herangezogen. Der NCV-Index ist normalverteilt mit einem Mittelwert von 0 und einer SD von 0,5. Je höher das Ausmaß der Demyelinisierung, desto kleiner der Index-Wert.

Abkürzungen: ARSA: Arylsulfatase A; BSC: Best Supportive Care; eCRF: electronic Case Report Form; EJ: Early Juvenile; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; IDS: Integrierter Datensatz; ITT: Intention-to-Treat; k. A.: keine Angabe; LI: Late Infantile; LLOQ: Lower limit of quantification; MLD: Metachromatische Leukodystrophie; MRT: Magnetresonanztomographie; MSAS: Matched Sibling Analysis Set; n. a.: nicht anwendbar; NCV: Nerve Conduction Velocity; pU: pharmazeutischer Unternehmer; SD: Standardabweichung; - nicht berechenbar.

Tabelle 26: Angaben zur Exposition mit der Studienmedikation; Geschwisterkinderanalyse IDS vs. TIGET-NHx, MSAS

MSAS Exposition mit der Studienmedikation	OTL-200 N = 17	TIGET-NHx N = 15
Busulfan-Konditionierungsschema, n (%)		
SMAC	9 (53)	n. a.
MAC	8 (47)	
AUC-Gesamtwert von Busulfan (µg x h/L)		
Mittelwert (SD)	77785,2 (8067,6)	n. a.
Median (min; max)	78574,2 (63420; 87940)	
OTL-200		
Anzahl an CD34+-Zellen (× 10 ⁶ /kg)		
Mittelwert (SD)	12,5 (5,73)	n. a.
Median (min; max)	12,5 (4,20; 25,9)	

Abkürzungen: AUC: Area Under the Curve; IDS: Integrierter Datensatz; MAC: Myeloablative Konditionierung; MSAS: Matched Sibling Analysis Set; n. a.: nicht anwendbar; SD: Standardabweichung; SMAC: Sub-myeloablative Konditionierung.

Mortalität

Die Ereigniszeitanalyse zum Alter zum Zeitpunkt des Todes kann nicht nachvollzogen werden (siehe Tabelle 27). Bei der Operationalisierung des Überlebens von Geburt bis zum Tod jeglicher Ursache ist nicht ersichtlich, wie lange die Teilnehmenden jeweils nach Behandlung bzw. Matchingbaseline überlebten. Falls bspw. eine Person im Alter von 6 Jahren eingeschlossen wird und zeitnah verstirbt, wird für diese Person dennoch eine Überlebensdauer von 6 Jahren angegeben, obwohl der Großteil von diesem Zeitraum vor Studien- und Behandlungsbeginn lag. Die Auswertung wird aus Transparenzgründen dargestellt. Eine Überlebensdauer ab Matchingbaseline, die zur Interpretation wichtig wäre, wurde jedoch nicht vorgelegt.

Aufgrund der vorliegenden Angaben wird zudem von einem Immortal-Time-Bias ausgegangen, der möglicherweise beide Arme betrifft. Demnach konnten die 3 verstorbenen Personen im IDS grundsätzlich nicht für das MSAS berücksichtigt werden, da das Matching auf dem Alter der behandelten Personen 2 Jahre nach der Gentherapie basierte – die 3 Personen sind jedoch alle vor Erreichen von Jahr 2 nach der Gentherapie verstorben (siehe Kapitel 3.2 Mortalität). Es ist unklar, ob diese 3 Patientinnen und Patienten Geschwisterkinder in der TIGET-NHx-Studie hatten und andernfalls für das Matching in Frage gekommen wären.

Tabelle 27: Ergebnisse für „Alter zum Zeitpunkt des Todes“ zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258), 26.05.2025 (HE 205029) und 01.11.2021 (TIGET-NHx); MSAS

MSAS Alter zum Zeitpunkt des Todes	OTL-200 N = 17	TIGET-NHx N = 15
Tod, n (%)	0 (0)	8 (53)
Zensierungen ¹⁾ , n (%)	17 (100)	7 (47)
Zensierungsgründe	k. A.	k. A.
Beobachtungsdauer ²⁾ (in Jahren), Median (min; max)	10,6 (7,64; 15,2)	4,54 (1,96; 17,9)
Überlebensdauer ab Matchingbaseline	k. A.	k. A.
Überlebensdauer ab Geburt ³⁾ (in Jahren), Median [95%-KI]	- [-; -]	10,2 [5,2; -]
Hazard Ratio [95%-KI]; p-Wert ⁴⁾	-	

¹⁾ Gemäß Modul 4 wurde für den Endpunkt Gesamtüberleben der Tod als Ereignis definiert; Personen ohne Ereignis wurden zum letzten verfügbaren Studienbesuch rechtszensiert.

²⁾ Gemäß SAP ist die Beobachtungsdauer für mit OTL-200 behandelte Kinder definiert als die Zeit von der Gentherapie bis zum Ende der letzten Studienvsiste. Für Teilnehmende der TIGET-NHx-Studie ist die Beobachtungsdauer definiert als der Zeitraum vom ersten Kontakt bis zum Abschluss der letzten Studienvsiste.

³⁾ Das Gesamtüberleben ist in den IDS-Studien operationalisiert als Zeitraum von Geburt bis zum Tod aus jeglicher Ursache. Die Überlebensdauer spiegelt die Überlebensdauer der verstorbenen Teilnehmenden wider.

⁴⁾ Das Hazard Ratio, das 95%-KI und der p-Wert sollten gemäß Modul 4 post-hoc mittels Cox-Regression mit Behandlung (OTL-200, TIGET-NHx) als kategoriale Variable bestimmt werden.

Abkürzungen: IDS: Integrierter Datensatz; k. A.: keine Angabe; KI: Konfidenzintervall; MSAS: Matched Sibling Analysis Set; SAP: statistischer Analyseplan; - : nicht berechenbar.

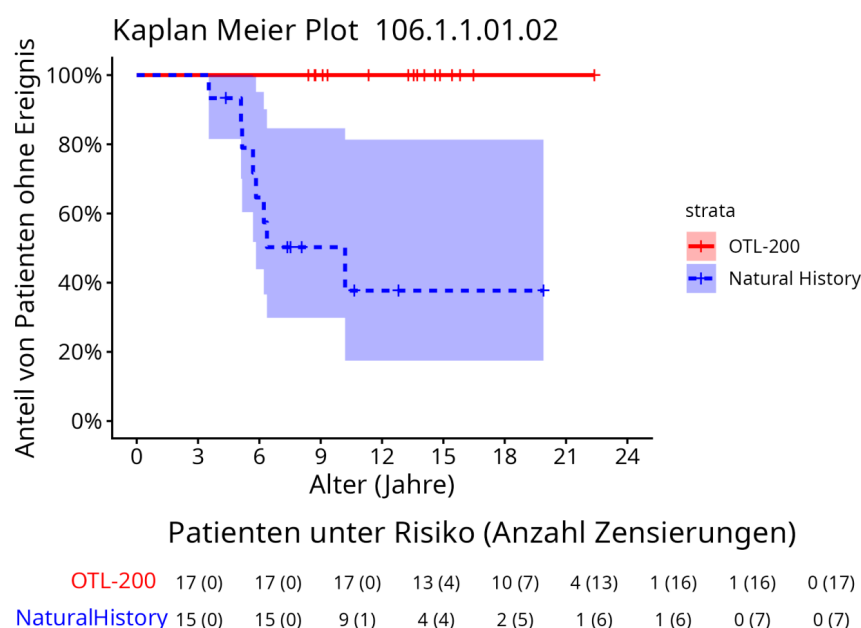


Abbildung 4: Kaplan-Meier-Kurven zu „Alter zum Zeitpunkt des Todes“ im IDS und in der Studie TIGET-NHx; MSAS [27]

Endpunktunabhängig wird in Modul 4 eine mediane Beobachtungsdauer von 10,6 Jahren (min; max: 7,64; 15,2) für die behandelten Kinder sowie von 4,54 (min; max: 1,96; 17,9) für die unbehandelte Vergleichsgruppe berichtet. Der Endpunkt „Gesamtüberleben“ ist jedoch operationalisiert als Zeitraum von Geburt bis zum Tod aus jeglicher Ursache, während die

Beobachtungsdauer für mit OTL-200 behandelte Kinder definiert ist als die Zeit von der Gentherapie bis zum Ende der letzten Studienvisite und für Teilnehmende der TIGET-NHx-Studie als der Zeitraum vom ersten Kontakt bis zum Abschluss der letzten Studienvisite.

Den Studienunterlagen kann entnommen werden, dass 4 Kinder der ITT-Population des IDS die Studienteilnahme vorzeitig abbrachen, darunter 3 Kinder, die verstarben, und 1 Kind, das als „Lost-to-Follow-up“ kategorisiert wurde. Ob letzteres im MSAS enthalten ist, ist nicht ersichtlich. Jedoch lässt sich aus der geringen Anzahl an Studienabbrüchen unter den Überlebenden ableiten (siehe Tabelle 16), dass entweder alle Zensierungen oder zumindest 16 der 17 Zensierungen darauf zurückzuführen sind, dass zum Studienende das Ereignis nicht eingetreten war. Die Zensierungsgründe in der unbehandelten Vergleichsgruppe lassen sich jedoch nicht entsprechend nachvollziehen. Gemäß den nachgereichten Post-hoc-Analysen wurden für die ITT-Population der Studie TIGET-NHx (N = 43) 13 Personen für die Ereigniszeitanalyse zensiert, da sie noch Teil der laufenden Studie und zum Zeitpunkt des Datenschnitts noch nicht verstorben waren, während 9 Personen zensiert wurden, weil sie die Studie vorzeitig abgebrochen hatten. Wie viele der betreffenden Personen aus der ITT-Population, die die Studie vorzeitig abbrachen, auch im vorliegenden MSAS (N = 15) enthalten sind, kann anhand der vorliegenden Informationen nicht nachvollzogen werden. Somit kann nicht abschließend beurteilt werden, ob informative Zensierungen vorlagen.

Morbidität

GMFC-MLD

Analog zur Analyse „Alter zum Zeitpunkt des Todes“ kann die Analyse „Alter bis zum GMFC-MLD Level ≥ 5 “ nicht nachvollzogen werden und wird nur aus Transparenzgründen dargestellt. Eine Auswertung zur Dauer bis zum Ereigniseintritt nach Matchingbaseline liegt nicht vor. Darüber hinaus liegen für die vorliegende Ereigniszeitanalyse Unklarheiten zur Imputation bei fehlenden Werten vor.

Zudem ist nicht ersichtlich, wie mit Respondern zu Matchingbaseline umgegangen wurde. Aufgrund der Operationalisierung „Alter bis zum GMFC-MLD Level ≥ 5 “ wird davon ausgegangen, dass etwaige Responder zu Matchingbaseline als Responder gezählt wurden. Wie viele Personen davon betroffen sein konnten, kann vorliegend nicht beziffert werden: es fehlen eine eindeutige Matchingbaseline-Definition für gematchte Geschwisterkinder aus der Studie TIGET-NHx und Angaben dazu, welche GMFC-MLD-Level diese zu Matchingbaseline aufwiesen.

Tabelle 28: Ergebnisse für „Anteil der Personen mit einem GMFC-MLD-Level ≥ 5 zum letzten Follow-up“ und „Alter bis zum GMFC-MLD-Level ≥ 5 “ zu den Datenschnitten 30.06.2025 (Studie 201222), 30.06.2025 (CUP 207394), 26.05.2025 (CUP 206258), 26.05.2025 (HE 205029) und 01.11.2021 (TIGET-NHx); MSAS

MSAS GMFC-MLD ¹⁾	OTL-200 N = 17	TIGET-NHx N = 15
Anteil der Personen mit einem GMFC-MLD-Level ≥ 5 zum letzten Follow-up²⁾		
GMFC-MLD-Level ≥ 5 , n (%) ³⁾		
Ja	2 (12)	14 (93)
Nein	15 (88)	1 (7)
Alter bis zum GMFC-MLD-Level $\geq 5$²⁾		
Zensierungen ⁴⁾ , n (%)	15 (88)	1 (7)
Zensierungsgründe	k. A.	k. A.
Beobachtungsdauer ⁵⁾ (in Jahren), Median (min; max)	10,6 (7,64; 15,2)	4,54 (1,96; 17,9)

MSAS GMFC-MLD ¹⁾	OTL-200 N = 17	TIGET-NHx N = 15
Zeit bis zum GMFC-MLD-Level ≥ 5 ab Matchingbaseline	k. A.	k. A.
Alter (in Jahren), Median [95%-KI]	- [13,2; -]	3,62 [2,67; 6,32]
Hazard Ratio [95%-KI]; p-Wert ⁶⁾	0,02 [0; 0,19]; < 0,0001	

¹⁾ Klassifikationssystem zur Beschreibung grobmotorischer Funktionen von Kindern mit MLD, die mindestens 18 Monate alt sind. Das Klassifikationssystem umfasst 7 verschiedene Stufen. GMFC-MLD-Level ≥ 5 : Verlust der Fähigkeiten, sich selbstständig fortbewegen und selbstständig sitzen zu können, Kopfkontrolle erhalten.

²⁾ Post-hoc-Analyse.

³⁾ Gemäß Modul 4 soll zwar bei fehlenden Werten eine „Worst-Case-Imputation“ durchgeführt werden, bei welcher ein fehlender Wert als Non-Response gewertet wird. Dabei ist zum einen unklar, ob dies für alle Arten von fehlenden Werten (inkl. aufgrund von Todesfällen) galt. Zum anderen kann dies jedoch insbesondere inhaltlich nicht nachvollzogen werden, da die Angaben im Zuge der Auswertungen implizieren, dass eine Response ab einem Level 5 vorlag und höhere Werte beim GMFC-MLD eine größere Einschränkung der grobmotorischen Fähigkeiten bedeuten. Demnach entspräche die Wertung von fehlenden Werten als Response einer „Worst-Case-Imputation“ und nicht – wie in Modul 4 beschrieben – die Wertung von fehlenden Werten als Non-Response. Es ist demnach unklar, wie imputiert wurde.

⁴⁾ Insofern zum Datenschnitt kein Ereignis vorlag, wurden Patientinnen und Patienten zum Zeitpunkt der letzten GMFC-MLD-Erhebung zensiert.

⁵⁾ Gemäß SAP ist die Beobachtungsdauer für mit OTL-200 behandelte Kinder definiert als die Zeit von der Gentherapie bis zum Ende der letzten Studienvisite. Für Teilnehmende der NHx-Studie ist die Beobachtungsdauer definiert als der Zeitraum vom ersten Kontakt bis zum Abschluss des letzten Studienvisite.

⁶⁾ Das Hazard Ratio, das 95%-KI und der p-Wert wurden post-hoc mittels Cox-Regression mit Behandlung (OTL-200, TIGET-NHx) als kategoriale Variable bestimmt.

Abkürzungen: GMFC-MLD: Gross Motor Function Classification in MLD; k. A.: keine Angabe; KI: Konfidenzintervall; MLD: Metachromatische Leukodystrophie; MSAS: Matched Sibling Analysis Set; SAP: statistischer Analyseplan; - : nicht berechenbar.

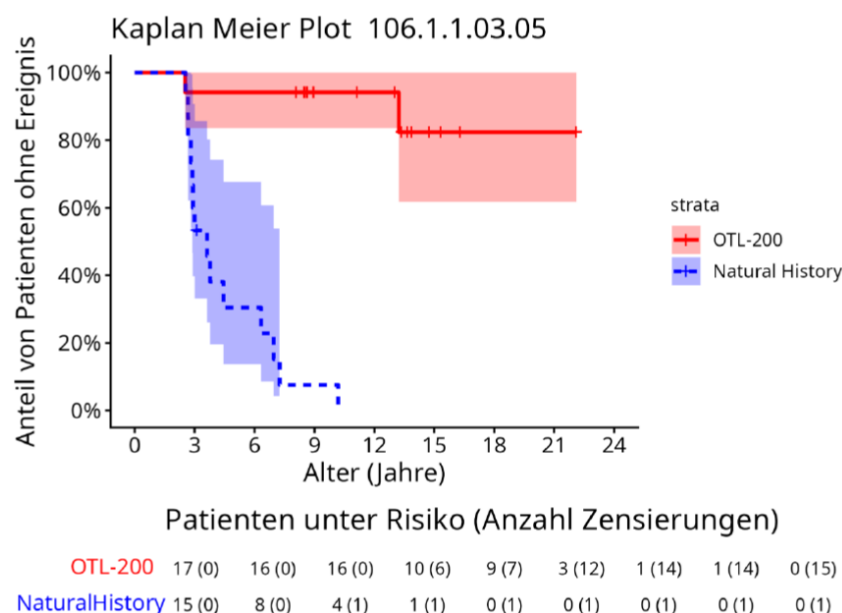


Abbildung 5: Kaplan-Meier-Kurven zum „Alter bis zum GMFC-MLD-Level ≥ 5 “ im IDS und in der Studie TIGET-NHx; MSAS [27]

Für den Endpunkt GMFC-MLD-Level ≥ 5 bestehen mehrere Unsicherheiten hinsichtlich der Datengrundlage und der statistischen Auswertung. Insbesondere ist (analog zu den einarmigen Daten) nicht hinreichend beschrieben, wie mit fehlenden GMFC-MLD-Werten

umgegangen wurde; entsprechende Angaben in den vorliegenden Studienunterlagen fehlen und die Angaben in Modul 4 sind widersprüchlich (siehe Tabelle 23).

Für das historische Kontrollkollektiv der TIGET-NHx-Studie erfolgte die Auswertung laut Modul 4 anhand von Längsschnittdaten. Es bleibt jedoch unklar, in welchem Umfang hierfür prospektiv bzw. retrospektiv erhobene Daten oder eine Kombination aus beiden Datenquellen verwendet wurde und welchen Anteil retrospektive Daten an der Analyse hatten. Auch die Beobachtungsdauer ab Matchingbaseline lässt sich für diesen Endpunkt nicht abschließend beurteilen. Zwar werden die Anzahl der Zensierungen und im Kaplan-Meier-Plot die Verläufe bis zum jeweiligen Ereignis bzw. Ausscheiden aus der Risikomenge dargestellt, Angaben zur Beobachtungsdauer ab Matchingbaseline sowie zu den Gründen der Zensierungen liegen jedoch nicht vor. Während im OTL-200-Arm 15 von 17 Personen (88 %) zensiert wurden, war dies im TIGET-NHx-Arm lediglich bei 1 von 15 Personen (7 %) der Fall.

Darüber hinaus war die Proportional-Hazards-Annahme mit einem p-Wert von 0,01 nicht erfüllt. Das berichtete Hazard Ratio von 0,02 [95%-KI: 0; 0,19] kann daher nicht als über die gesamte Beobachtungszeit konstanter relativer Behandlungseffekt interpretiert werden, sondern allenfalls als ein über den Zeitverlauf gemitteltes Effektmaß. Dies ist insbesondere deshalb relevant, weil die Anzahl der noch unter Risiko stehenden Personen im TIGET-NHx-Arm bereits früh deutlich abnimmt. Ab einem Alter von etwa 6 Jahren stehen nur noch sehr wenige Personen des historischen Kontrollarms für den Vergleich zur Verfügung; mit 9 Jahren befindet sich lediglich noch eine Person unter Risiko und zu späteren Zeitpunkten keiner mehr. Inwiefern fehlende Werte und Imputationen dies beeinflussten, ist unklar. Aussagen zu einem vergleichenden Behandlungseffekt in höheren Altersbereichen sind nur sehr eingeschränkt bzw. nicht mehr belastbar. Auch davor sind jedoch die wesentlichen Unklarheiten zu den Imputationen zu berücksichtigen.

Der beim Endpunkt „Gesamtüberleben“ beschriebene Immortal-Time-Bias im Zusammenhang mit dem Matching-Verfahren sowie der durch das Matching-Vorgehen möglicherweise bedingte Ausschluss der 3 im IDS verstorbenen Personen ist auch beim Endpunkt GMFC-MLD zu berücksichtigen.

GMFM

Für die gematchten Teilnehmenden der TIGET-NHx-Studie lässt sich die Vollständigkeit der GMFM-Datenerhebung anhand der vorliegenden Angaben nicht hinreichend beurteilen. Insbesondere können keine belastbaren Rücklaufquoten bestimmt werden, da für die einzelnen Follow-up-Zeitpunkte nicht angegeben ist, wie viele Patientinnen und Patienten zu diesem Zeitpunkt noch lebten und damit grundsätzlich für eine Erhebung zur Verfügung standen. Da die Rücklaufquoten bei Berechnung ohne Berücksichtigung der Todesfälle bereits zum ersten berichteten Follow-up-Zeitpunkt zu Jahr 2 nach (insbesondere für die gematchten Personen der TIGET-NHx-Studie nicht näher definierter) Baseline unter 70 % lagen, werden die Ergebnisse des GMFM im Rahmen der Geschwisterkinderanalyse nicht dargestellt. Es ist außerdem nicht nachvollziehbar, dass im gematchten TIGET-NHx-Arm für die Dimensionen A–E zu Jahr 2 eine Rücklaufquote von nur 33 % vorliegt, während für den Gesamtscore zu Jahr 2 insgesamt 60 % im TIGET-NHx-Arm Werte vorliegen hatten. Da sich der Gesamtscore aus dem Mittel der Dimensionen berechnet, sollten sich die Rücklaufquoten zwischen Gesamtscore und Dimensionen nicht unterscheiden.

Zusätzlich bestehen Unsicherheiten hinsichtlich der Erhebung der GMFM-Daten im historischen Kontrollarm. Für die TIGET-NHx-Studie wurden neben prospektiv erhobenen Daten auch retrospektive Daten herangezogen. Der Anteil retrospektiver Daten ist jedoch nicht angegeben. Auch unter Berücksichtigung der im Stellungnahmeverfahren der Vorbewertung [12] getätigten Erklärungen des pU bleibt unklar, ob und auf welche Weise

GMFM-Werte retrospektiv valide erhoben bzw. rekonstruiert werden konnten. Der GMFM basiert auf der standardisierten direkten Beobachtung der Fähigkeit eines Kindes, 88 vorgegebene motorische Aufgaben auszuführen, sodass eine retrospektive Ermittlung nur möglich erscheint, wenn eine entsprechende standardisierte Erhebung in der Vergangenheit stattfand und mittels Videoaufnahme festgehalten wurde. Zusätzlich bleibt unklar, weshalb im TIGET-NHx-Arm zu Jahr 2 für mehrere Personen kein GMFM-Wert vorliegt, obwohl das Matching für den indirekten Vergleich anhand der MLD-Form und des Alters zu Jahr 2 nach der Gentherapie erfolgte und beim Matching der erste vollständige Datensatz berücksichtigt werden sollte.

Indirekter Vergleich fröhsymptomatische Patientinnen und Patienten

Tabelle 29: Charakterisierung der Studienpopulation; indirekter Vergleich fröhsymptomatische Patientinnen und Patienten des IDS vs. TIGET-NHx, ITT

IDS vs. TIGET-NHx, ITT Charakterisierung der Studienpopulation	OTL-200 N = 9	TIGET-NHx N = 43
<i>Alter zu Krankheitsbeginn¹⁾ (Monate)</i> Mittelwert (SD) Median (min; max)	56,6 (17,1) 60 (26; 83)	27,9 (18,6) 19 (9; 75)
<i>Alter zu Diagnosestellung (Monate)</i> Mittelwert (SD) Median (min; max)	68,2 (31,6) 68,1 (20,2; 131,9)	41,3 (18,5) 33,9 (18,6; 91,4)
<i>Alter zu Screening/Studieneinschluss (Monate)</i> Mittelwert (SD) Median (min; max)	70,9 (32,6) 69,8 (21,6; 136,2)	90,5 (47,2) 81,6 (30,7; 214,6)
<i>Alter zum Zeitpunkt des 1. Kontaktes²⁾ (Monate)</i> Mittelwert (SD) Median (min; max)	71 (32,6) 69,9 (21,6; 136,4)	32 (18,2) 23,7 (14,5; 74,1)
<i>Alter zum Zeitpunkt der Genterapie (Monate)</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	8 (88,9) ⁴⁾ 71,7 (35,2) 70 (23,3; 139,7)	k. A.
<i>Alter bei letzter Untersuchung (Monate)</i> Mittelwert (SD) Median (min; max)	15,3 (5,54) 17 (6,56; 22,4)	9,47 (5,3) 8,07 (2,70; 25,3)
<i>Geschlecht, n (%)</i> Weiblich Männlich	6 (67) 3 (33)	23 (53) 20 (47)
<i>Abstammung (genetisch), n (%)</i> Weiß - kaukasisch/europäisch Weiß - arabisch/nordafrikanische Herkunft	8 (89) 1 (11)	39 (91) 4 (9)
<i>Ethnie, n (%)</i> Nicht Hispanisch oder Latino Hispanisch oder Latino	9 (100) 0	41 (95) 2 (5)
<i>MLD-Form, n (%)</i> Spätfantil (LI) Frühjuvenil (EJ) ⁵⁾	1 (11) 8 (89)	28 (65) ⁶⁾ 15 (35)
<i>Symptomstatus zu Screening/Studieneinschluss, n (%)</i> Präsymptomatisch Fröhsymptomatisch	1 (11) 8 (89)	0 43 (100)
<i>Symptomstatus zum Zeitpunkt der Genterapie, n (%)</i> Präsymptomatisch Fröhsymptomatisch	k. A.	k. A.
<i>GMFM-Score (%)⁷⁾ zu Baseline⁸⁾</i> n (%) ³⁾ Mittelwert (SD) Median (min; max)	9 (100) ⁴⁾ 84,6 (9,67) 86,6 (71,1; 98,7)	37 (86) ⁴⁾ 16,6 (25,5) 4,53 (0,78; 96,4)

IDS vs. TIGET-NHx, ITT Charakterisierung der Studienpopulation	OTL-200 N = 9	TIGET-NHx N = 43
GMFC-MLD⁹⁾ zu Baseline⁸⁾, n (%)		
Stufe 0	2 (22)	0
Stufe 1	7 (78)	12 (28)
Stufe 2	0	11 (26)
Stufe 3	0	0
Stufe 4	0	2 (5)
Stufe 5	0	6 (14)
Stufe 6	0	12 (28)
k. A. ¹⁰⁾	0	0
ARSA-Aktivität_{PBMC} (nmol/mg/h) zu Baseline⁸⁾		
n (%) ³⁾	9 (100) ⁴⁾	1 (2,33) ⁴⁾
Mittelwert (SD)	26 (0,73)	25,8 (-)
Median (min; max)	25,8 (25,8; 28)	25,8 (25,8; 25,8)
MRT-Score¹¹⁾ zu Baseline⁸⁾		
n (%) ³⁾	9 (100) ⁴⁾	36 (83,7) ⁴⁾
Mittelwert (SD)	7,92 (4,32)	15,2 (8,22)
Median (min; max)	10 (0,50; 12)	13,4 (1; 28,5)
NCV-Index¹²⁾ zu Baseline⁸⁾		
n (%) ³⁾	9 (100) ⁴⁾	29 (67,4) ⁴⁾
Mittelwert (SD)	-7,01 (2,61)	-9,65 (4,69)
Median (min; max)	-7,93 (-9,45; -3,11)	-11 (-14,4; 0)
Begleitende Therapie zu Studienbeginn, n (%)		
Ja	k. A.	k. A.
Nein		
Vorthherapie zu Studienbeginn, n (%)		
Ja	k. A.	k. A.
Nein		

¹⁾ Bei präsymptomatischen Patientinnen und Patienten wurde das vorhergesagte Alter zu Krankheitsbeginn auf der Grundlage des Alters beim Auftreten der ersten Symptome bei den älteren Geschwistern der Patientin / des Patienten berechnet.

²⁾ Für Kinder des OTL-200-Arms entspricht dies dem Alter zum Zeitpunkt der Baseline-Erhebung von GMFC-MLD, GMFM, NCV und MRT. Es ist unklar, ob dies auch für den TIGET-NHx-Arm zutrifft.

³⁾ Eine Angabe von n (%) erfolgt nur bei Charakteristika, für die keine vollständigen Daten vorlagen.

⁴⁾ Prozentzahlen aus Eigenberechnung.

⁵⁾ Für die Analysen wurde eine Person der Studie 201222 mit einer intermediären Form der frühjuvenilen Form (EJ) zugeordnet. Ob diese Person prä- oder frühsymptomatisch war, ist unklar.

⁶⁾ Da alle Personen im TIGET-NHx Arm frühsymptomatisch waren, entspricht diese Population nicht dem zugelassenen Anwendungsgebiet von OTL-200.

⁷⁾ Der Wertebereich der GMFM reicht von 0–100 %, wobei ein höherer Wert stärker ausgeprägten grobmotorischen Funktionen entspricht.

⁸⁾ Angabe „zu Baseline“ gemäß Modul 4. Es ist unklar, ob es sich um die jeweilige Baseline-Erhebung in den IDS-Studien bzw. der TIGET-NHx-Studie handelt. Für die TIGET-NHx-Studie liegt zudem keine Baseline-Definition vor.

⁹⁾ Klassifikationssystem zur Beschreibung grobmotorischer Funktionen von Kindern mit MLD, die mindestens 18 Monate alt sind. Das Klassifikationssystem umfasst 7 verschiedene Stufen (Stufe 0: altersentsprechendes, freies Gehen ohne Unterstützung; Stufe 6: Verlust jeglicher Fortbewegungsfähigkeit sowie jeglicher Kopf- und Rumpfkontrolle).

¹⁰⁾ Da das GMFC-MLD-Level auf der Fähigkeit zu Gehen basiert, kann es in der Regel erst ab einem Alter von mindestens 18 Monaten bestimmt werden. Somit konnten für Kinder mit einem Alter von < 18 Monate keine Baseline-Werte erhoben werden. Bei Kindern, die jünger sind als 18 Monate, aber schon laufen können, kann das GMFC-MLD-Klassifikationssystem gemäß pU jedoch angewendet werden. Ob alle fehlenden Werte darauf

zurückzuführen sind, dass die entsprechenden Kinder < 18 Monate alt waren bzw. noch nicht laufen konnten, kann den Studienunterlagen nicht entnommen werden.

¹¹⁾ Der MRT-Score wurde anhand einer adaptierten Loes' Skala ermittelt, die Subskalen zu Demyelinisierung (0–16 Punkte), b) globaler (0–4 Punkte) und fokaler (0–10,5 Punkte) Hirnatrophie sowie getriggertem Muster ('tigroid pattern') (0–1 Punkte) umfasst. Der Gesamtscore kann Werte von 0 (unauffällig) bis 31,5 (merklich verändert) annehmen. Ein Wert von > 0 gilt als abnormal.

¹²⁾ Der NCV-Index berechnet sich aus dem Durchschnitt von 4 z-Scores, die auf Testungen der Nervenleitgeschwindigkeit des Nervus suralis, Nervus medianus, Nervus fibularis profundus und Nervus ulnaris beruhen. Für die Berechnung der z-Scores wurden, je nach Alter der Patientinnen und Patienten, 3 altersbezogene Normkollektive herangezogen. Der NCV-Index ist normalverteilt mit einem Mittelwert von 0 und einer SD von 0,5. Je höher das Ausmaß der Demyelinisierung, desto kleiner der Index-Wert.

Abkürzungen: ARSA: Arylsulfatase A; EJ: Early Juvenile; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; IDS: Integrierter Datensatz; ITT: Intention-to-Treat; k. A.: keine Angabe; LI: Late Infantile; MLD: Metachromatische Leukodystrophie; MRT: Magnetresonanztomographie; NCV: Nerve Conduction Velocity; pU: pharmazeutischer Unternehmer; SD: Standardabweichung; - nicht berechenbar.