

Nutzenbewertung



**Gemeinsamer
Bundesausschuss**

**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: Voretigen Neparvovec

Datum der Veröffentlichung: 15. Juli 2019

Inhaltsverzeichnis

Tabellenverzeichnis.....	3
Abbildungsverzeichnis.....	4
Abkürzungsverzeichnis	5
Hintergrund.....	6
1 Einführung	7
2 Nutzenbewertung	8
2.1 Fragestellung.....	8
2.2 Zulassungsbegründende Studie.....	8
2.3 Liste der verwendeten Quellen.....	9
2.4 Studiencharakteristika	9
2.5 Endpunkte und statistische Methoden	15
2.5.1 Patientenrelevanz der Endpunkte.....	15
2.5.2 Charakterisierung der patientenrelevanten und ergänzend dargestellten Endpunkte	16
2.5.3 Verzerrungspotential auf Studien und Endpunktebene.....	34
2.5.4 Statistische Methoden.....	36
2.6 Ergebnisse zum Zusatznutzen	38
2.6.1 Charakterisierung der Studienpopulation	38
2.6.2 Mortalität.....	43
2.6.3 Morbidität.....	43
2.6.4 Lebensqualität.....	44
2.6.5 Sicherheit	44
3 Methodische Anmerkungen und Bewertung der Unterlagen.....	49
3.1 Zulassungsstatus und Zulassungspopulation von Studie 301.....	49
3.2 Design und Methodik der Studie	49
3.3 Wirksamkeit	50
3.4 Lebensqualität	52
3.5 Sicherheit.....	52
4 Anforderung an eine qualitätsgesicherte Anwendung	54
5 Zusammenfassung der Bewertung	55
Referenzen	57

Tabellenverzeichnis

Tabelle 1:	Charakterisierung der Studie 301	9
Tabelle 2:	Charakterisierung der Intervention in der Studie 301	14
Tabelle 3:	Bewertung der Patientenrelevanz der Endpunkte der Studie 301	15
Tabelle 4:	Beschreibung der Endpunkte der Kategorie Mortalität	16
Tabelle 5:	Beschreibung der Endpunkte der Kategorie Morbidität	17
Tabelle 6:	Umrechnungstabelle für den Mobilitätsscore.....	19
Tabelle 7:	Beschreibung der Endpunkte der Kategorie Lebensqualität	31
Tabelle 8:	Beschreibung der Endpunkte der Kategorie Sicherheit.....	32
Tabelle 9:	Erhebungszeitpunkte der berücksichtigten Endpunkte in der Studie 301	34
Tabelle 10:	Verzerrungspotential der Studie 301	34
Tabelle 11:	Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie 301	35
Tabelle 12:	Allgemeine Angaben zur Studie 301	39
Tabelle 13:	Charakterisierung der Studienpopulation der Studie 301 (ITT-Population).....	41
Tabelle 14:	Änderung im MLMT bilateral gemessen, Jahr 1B/C im Vergleich zu Baseline (ITT-Population)	43
Tabelle 15:	Ergebnisse der Nebenwirkungen – Zusammenfassung der UE zu Jahr 1 nach Baseline in der Studie 301 (Sicherheitspopulation)	45
Tabelle 16:	UE mit Inzidenz ≥ 10 % der Sicherheitspopulation in der Studie 301 innerhalb eines Jahres nach Baseline	45
Tabelle 17:	SUE mit Inzidenz ≥ 5 % in der Studie 301 innerhalb eines Jahres nach Baseline (Sicherheitspopulation)	47
Tabelle 18:	Schwere UE mit Inzidenz ≥ 5 % in der Studie 301 innerhalb eines Jahres nach Baseline (Sicherheitspopulation)	48
Tabelle 19:	Zusammenfassende Darstellung der Ergebnisse der Studie 301.....	55

Abbildungsverzeichnis

Abbildung 1: Studiendesign der Studie 301	13
Abbildung 2: Beispiel für einen der 12 Hindernisparcours des MLMT [10].....	17
Abbildung 3: Verwendete Beleuchtungsstärken beim MLMT der Studie 301	18

Abkürzungsverzeichnis

AAV	Adeno-assoziiertes virales Vektor
CHOP	Children's Hospital of Philadelphia
EMA	European Medicines Agency
EPAR	European Public Assessment Report
FDA	U.S. Food and Drug Administration
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
IOWA	University of Iowa
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
LCA	Lebersche kongenitale Amaurose
MID	Minimal Important Difference
mITT	modified Intention-to-Treat
MLMT	Multi-Luminanz-Mobilitätstest
MW	Mittelwert
PP	Per Protocol
pU	pharmazeutischer Unternehmer
RPE65	retinales Pigmentepithel-spezifisches 65 kDa-Protein
SAP	Statistischer Analyseplan
SD	Standardabweichung
SGB	Sozialgesetzbuch
SOP	Standard Operating Procedure
SUE	Schwerwiegende/s unerwünschte/s Ereignis/se
UE	Unerwünschte/s Ereignis/se
VerfO	Verfahrensordnung des G-BA
Vg	Vektorgenome

Hintergrund

Voretigen Neparvec 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 (G-BA), ob er die Nutzenbewertung selbst durchführt oder das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) beauftragt. Ausgehend von der gesetzlichen Vorgabe in § 35a Absatz 1 Satz 11 1. Halbs. SGB V, dass der Zusatznutzen eines Orphan Drug durch die Zulassung als belegt gilt, hat der G-BA in seiner Sitzung vom 15. März 2012 das Verfahren der Nutzenbewertung von Orphan Drugs dahingehend modifiziert, dass bei Orphan Drugs zunächst keine eigenständige Festlegung einer zweckmäßigen Vergleichstherapie mehr durch den G-BA als Grundlage der insoweit allein rechtlich zulässigen Bewertung des Ausmaßes eines gesetzlich zu unterstellenden Zusatznutzens erfolgt. Vielmehr wird ausschließlich auf der Grundlage der Zulassungsstudien das Ausmaß des Zusatznutzens durch den G-BA bewertet.

Dementsprechend hat der G-BA in seiner Sitzung am 15. März 2012 den mit Beschluss vom 1. August 2011 erteilten Auftrag an das IQWiG zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen gemäß § 35a Absatz 2 SGB V in der Weise abgeändert, dass bei Orphan Drugs eine Beauftragung des IQWiG mit der Durchführung einer Nutzenbewertung bei zuvor festgelegter Vergleichstherapie erst dann erfolgt, wenn der Umsatz des betreffenden Arzneimittels die gesetzliche Grenze von 50 Millionen Euro überschritten hat und damit einer uneingeschränkten Nutzenbewertung unterliegt (vgl. § 35a Absatz 1 Satz 12 SGB V).

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

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Voretigen Neparvec in seiner Sitzung am 9. Juli 2019 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 9. April 2019 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-Patienten in der Zielpopulation

Die Nutzenbewertung wird am 15. Juli 2019 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1 Einführung

Luxturna dient der subretinalen Anwendung zur Behandlung von erwachsenen und pädiatrischen Patientinnen und Patienten mit Sehverlust aufgrund einer erblichen Netzhautdystrophie, die auf nachgewiesenen biallelischen *RPE65*-Mutationen beruht, und die über ausreichend lebensfähige Netzhautzellen verfügen. Gemäß EPAR (European Public Assessment Report) beinhaltet der Term „erbliche Netzhautdystrophie aufgrund von biallelischer *RPE65*-Mutation“ sowohl die Lebersche kongenitale Amaurose (LCA) als auch Retinitis pigmentosa, da es sich um unterschiedliche phenotypische Ausprägungen mit einer bestimmten genetischen Mutation handelt. [17]

Die *RPE65*-Mutation führt zum Verlust der Funktion der 11-cis-Retinase, es kommt zu einem voranschreitenden Verlust der Sehfähigkeit. Betroffen sind sowohl die Stäbchen (Nachtsehen) als auch die Zapfen (Farbsehen), wobei der Funktionsverlust bei Zapfen aufgrund von alternativem Stoffwechselweg weniger stark ausgeprägt ist. [6]

Die LCA vom Typ 2 tritt bei 1 von 80.000 Personen in der EU auf. 10 % der Fälle sind auf eine *RPE65*-Mutation zurückzuführen. Sie wird autosomal rezessiv vererbt. Symptome können bereits im Alter von 2–3 Monaten auftreten [6], die Prävalenz liegt bei 1–9/100.000 [14]. Es kommt zu Verlust der Sehfähigkeit, Verringerung des Sehfeldes, Nachtblindheit und Nystagmus. Es fällt den betroffenen Personen schwer, Aktivitäten des täglichen Lebens bereits bei Tageslicht zu verrichten und sie können bereits in der frühen Jugend erblinden. [6]

Retinitis pigmentosa tritt bei einem von 4.000 Individuen auf. Schätzungsweise ist die Erkrankung bei 3 % der Fälle auf eine *RPE65*-Mutation zurückzuführen [6], die Prävalenz ist 1–5/10.000 [8]. Im Gegensatz zur LCA vom Typ 2 ist das erste Auftreten der Erkrankung variabler und verläuft weniger stark progressiv. Die ersten Symptome treten erst ab einem Alter von 10 Jahren auf. Die Sehfähigkeit bleibt besser erhalten als bei der LCA vom Typ 2. [6]

Die Therapiemöglichkeiten beschränken sich auf die Verwendung von Hilfsmitteln, wie Sehhilfen, Lesebrillen oder Blindenlangstöcke. Sofern eine Lichtwahrnehmung noch vorhanden ist kann auch der Einsatz einer Netzhautprothese diskutiert werden. [6]

Voretigen Neparvovec ist ein Gentransfer-Vektor, der das Kapsid eines Adeno-assoziierten viralen Vektors (AAV) vom Serotyp 2 als Transportvehikel für die komplementäre DNA des humanen retinalen Pigmentepithel-spezifischen 65 kDa-Proteins (*RPE65*) in die Retina verwendet. Dies führt zur Transduktion von retinalen Pigmentepithelzellen mit komplementärer DNA, die für das normale humane *RPE65*-Protein kodiert, wodurch diese in der Lage sind das Enzym selbst zu exprimieren. Voretigen Neparvovec wird aus einem natürlich vorkommenden AAV mittels rekombinanter DNA-Techniken gewonnen. Die zu verabreichende Dosis pro Auge ist $1,5 \times 10^{11}$ Vektorgenome (Vg) in dem zu verabreichenden Volumen von 0,3 ml. [17]

2 Nutzenbewertung

2.1 Fragestellung

Voretigen Neparvovec (Luxturna®) ist ein Arzneimittel mit dem Status „Orphan Drug“. Gemäß § 35a Abs. 1 Satz 11 1. Halbs. SGB V gilt für Orphan Drugs der medizinische Zusatznutzen bereits durch die Zulassung als belegt. Der G-BA bestimmt bei Orphan Drugs, die einen Umsatz von 50 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. In die Bewertung werden gemäß Zulassungsstatus Personen in folgender Indikation eingeschlossen:

Erwachsene und pädiatrische Patienten mit Sehverlust aufgrund einer erblichen Netzhautdystrophie, die auf nachgewiesenen biallelischen *RPE65*-Mutationen beruht, und die über ausreichend lebensfähige Netzhautzellen verfügen

2.2 Zulassungsbegründende Studie

Pivotaler Studie 301

“A Safety and Efficacy Study in Subjects with Leber Congenital Amaurosis (LCA) Using Adeno-Associated Viral Vector to Deliver the Gene for Human *RPE65* to the Retinal Pigment Epithelium (RPE).”

Weitere vom pU eingereichte Studien

- Studie 101: “A Phase 1 Safety Study in Subjects with Leber Congenital Amaurosis (LCA) Using Adeno-Associated Viral Vector to Deliver the Gene for Human *RPE65* into the Retinal Pigment Epithelium (RPE).”
- Studie 102: “A Follow-On Study to Evaluate the Safety of Readministration of Adeno-Associated Viral Vector Containing the Gene for Human *RPE65* [AAV2- hRPE65v2] to the Contralateral Eye in Subjects with Leber Congenital Amaurosis (LCA) Previously Enrolled in a Phase 1 Study”
- Studie LTFU/302: “A Long-Term Follow-Up Study in Subjects Who Received an Adenovirus-Associated Viral Vector Serotype 2 Containing the Human *RPE65* Gene (AAV2-hRPE65v2) Administered via Subretinal Injection”

Für die Nutzenbewertung wird die pivotaler, randomisierte, kontrollierte, offene Phase-III-Studie 301 herangezogen. Die Studien 101, 102 und LTFU/302 werden nicht berücksichtigt. Bei Studie 101 handelt es sich um eine einarmige Dosisfindungsstudie, bei der Personen ab 8 Jahren mit einer Sehschärfe von $\leq 20/160$ bzw. Gesichtsfeld $< 20^\circ$ teilnehmen konnten. Die Patientinnen und Patienten dieser Studie konnten an der Follow-up-Studie 102 teilnehmen. Die Endpunkte ähnelten denen der Studie 301, wobei a priori kein primärer bzw. keine sekundären Effektivitätseindpunkte definiert wurden. In Studie 101 erhielten zunächst 3 der 12 in die Studie eingeschlossenen Personen in ein Auge eine Dosis von $1,5 \times 10^{11}$ Vg in einem Volumen von 300 μ l. Alle 12 Personen der Studie 101 wurden ein Jahr nachbeobachtet und erhielten in der darauf folgenden Studie 102 die zulassungsrelevante Dosierung in das noch nicht behandelte Auge. Die Nachbeobachtung der Studie 102 war ebenfalls für ein Jahr geplant. Aufgrund des einarmigen Studiendesigns, der geringen Anzahl an Personen, die die zulassungsrelevante Dosierung in beide Augen erhielt, und

dem langen zeitlichen Abstand zwischen den beiden Injektionen wird diese Studie nicht in der Nutzenbewertung dargestellt.

Bei der Studie LTFU/302 handelt es sich um eine Langzeit-Beobachtungsstudie von allen Personen, die die Prüfmedikation erhalten haben und die anschließend bis zu 15 Jahre nachbeobachtet werden sollen, wobei in den ersten 5 Jahren Studienvisiten und im Anschluss ein Nachverfolgen per Telefon geplant ist. Für die Personen der Studie 301 wurden nur die dort bereits definierten Wirksamkeits- und Sicherheitsendpunkte weitererhoben. Zum aktuellsten Datenschnitt vom 5. März 2017 lagen für die Interventionsgruppe Daten zu Jahr 3 nach Baseline (3 Jahre nach Administration der Prüfmedikation) sowie für die Kontrollgruppe zu Jahr 2 nach Injektion der Prüfmedikation zu den Effektivitäts- und Sicherheitsendpunkten vor. Die Studie LTFU/302 wird für die Nutzenbewertung nicht herangezogen, da keine vergleichenden Daten (Prüfpräparat vs. beobachtendes Abwarten) vorliegen, während in Studie 301 Daten für den Vergleich der Interventionsgruppe mit der Kontrollgruppe ohne Intervention zu einem Jahr existieren.

2.3 Liste der verwendeten Quellen

Zur Bestimmung des Ausmaßes des Zusatznutzens für Voretigen Neparvovec wurden folgende Studien und Daten herangezogen:

- Herstellerdossier zu Voretigen Neparvovec [16]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [6]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der Studie 301 [24,25,26]

2.4 Studiencharakteristika

Die Unterlagen zum Nachweis des Ausmaßes des Zusatznutzens für Voretigen Neparvovec basieren auf der Zulassungsstudie 301. Die Studie und die Intervention werden in den folgenden 2 Tabellen charakterisiert.

Tabelle 1: Charakterisierung der Studie 301

Charakteristikum	Studie 301
Design	<p>Randomisierte, kontrollierte, offene Phase-III-Studie im Parallelgruppenvergleich, bei der Voretigen Neparvovec als gentherapeutische Intervention in Form einer subretinalen Injektion verglichen wurde mit beobachtendem Abwarten in der Kontrollgruppe. Ein Wechsel der Kontrollgruppe in den Interventions-Arm war nach mindestens einem Jahr möglich. Die Beobachtungsdauer nach Injektion der Prüfmedikation ins 2. Auge betrug in beiden Gruppen 1–1,5 Jahre, wobei in der Kontrollgruppe die Möglichkeit der Behandlung mit dem Prüfpräparat erst nach Visite zu Jahr 1 bestand. Die Patientinnen und Patienten konnten im Anschluss in der Studie LTFU/302 bis zu 15 Jahre weiterbeobachtet werden. Das Studiendesign ist in Abbildung 1 dargestellt.</p> <p>Der primäre Endpunkt untersuchte die Änderung im Mult-Luminanz-Mobilitätstest (MLMT, kurz: „Mobilitätstest“) bilateral gemessen (unverdeckt mit beiden Augen) nach Jahr 1B (= Jahr 1 nach Behandlung des zweiten Auges in der Interventionsgruppe) bzw. Jahr 1C (= Jahr 1 nach der Baseline der Kontrollgruppe) zwischen den Behandlungsgruppen.</p>

Charakteristikum	Studie 301
	<p>Die Randomisierung erfolgte im Verhältnis 2:1 in Form einer Block-Randomisierung (Blockgröße von 3), stratifiziert nach Alter zum Zeitpunkt der Screeningvisite (≥ 10 vs. < 10 Jahre) und Ergebnislevel des Mobilitätstests zum Screening mit dem am schlechtesten im Test abscheidenden Auge (bestehen bei ≥ 125 vs. < 125 Lux).</p> <p>Für die Randomisierung war keine separate Studienvisite vorgesehen. Sie fand nach dem Screening statt und erfolgte durch ein unabhängiges Unternehmen. Die Bekanntgabe an das Studienpersonal fand zum Ende der Baseline-Visiten statt, welche innerhalb von 90 Tagen nach der Screening-Visite stattfinden sollte.</p> <p>Die Teilnehmenden der Interventionsgruppe konnten die Intervention nicht gleichzeitig in jedes Auge erhalten. Nach der Applikation ins erste Auge (Tag 0A; höchstens 90 Tage nach Baseline) erfolgte innerhalb von 12 ± 6 Tagen die Injektion in das zweite Auge (Tag 0B). Die Kontrollgruppe erhielt keine Injektion (Tag 0C = Baseline). Nach der Visite zu Jahr 1C der Kontrollgruppe konnten die Teilnehmenden, sofern die Ein-/Ausschlusskriterien noch erfüllt waren, die Prüfmedikation erhalten. Die Vorgaben zu den Injektionen in beide Augen war ähnlich der Interventionsgruppe.</p> <p>Im Protokoll konnten keine Kriterien zur Festlegung des ersten bzw. zweiten Auges identifiziert werden.</p> <p>Es war zu Beginn der Studie festgelegt, zunächst einen Abstand von mindestens 2 Wochen zwischen der Lieferung des Vektors zu den Studienteilnehmenden für die ersten 5 Betroffenen in jedem Studienzentrum einzuhalten. Anschließend konnte diese Staffelung, sofern vom DSMB genehmigt, aufgehoben werden. Dies erfolgte im Oktober 2013.</p> <p>Die Rekrutierung konnte gestoppt werden, sobald 18 Patientinnen und Patienten eine Injektion in das zweite Auge erhalten haben und mindestens 9 Personen in den Kontrollarm eingeschlossen wurden. Durch die Blockrandomisierung war eine Fallzahl von 31 Teilnehmenden möglich.</p> <p><u>Kriterien für Studienabbruch und Nachrandomisierung von Patientinnen/Patienten</u></p> <p>Ein Studienabbruch war für die teilnehmende Person jederzeit möglich. Behandlungsbedingte UE wurden in der Interventionsgruppe bis zu 30 Tage nach Studienabbruch erfasst. Nach Studienabbruch einer Patientin / eines Patienten bestand die Möglichkeit, eine neue Person nachzurekrutieren. In dem Fall erfolgte eine neue Randomisierung, d. h. unabhängig davon, ob die ausscheidende Person aus der Interventions- oder der Kontrollgruppe stammte.</p> <p>Das klinische Studienpersonal hatte die Möglichkeit die zweite Injektion nicht vorzunehmen, sofern chirurgische, medizinische oder psychologische Bedenken bestanden. Auch konnte der Patient / die Patientin die Injektion verweigern. Die Person nahm in dem Fall an allen Studienvisiten teil. Sofern keine Bedenken mehr bestanden und der Zeitraum für eine zweite Injektion mehr als 18 Tage zurücklag, konnte eine zweite Injektion ein Jahr nach Baseline erfolgen. Die Ergebnisse zu einem Jahr zur Wirksamkeit nach der unilateralen Applikation konnten für die Wirksamkeitsanalyse verwendet werden.</p> <p>Falls nur eine unilaterale Administration der Prüfmedikation stattfand, bestand ebenfalls die Möglichkeit, erneut eine Person in die Studie einzuschließen. Dies war bis zu zwei Mal während der Studie möglich. Sofern neue Personen für Teilnehmende, die nur eine unilaterale Applikation der Prüfmedikation erhalten haben, in die Studie aufgenommen wurden, wurden die Personen mit nur unilateraler Applikation nach Jahr 1 von der Effektivitätsanalysen ausgenommen.</p> <p>Die Studie ist abgeschlossen. Es liegt der finale Studienbericht (Amendment 003) vom 09.08.2017 vor.</p>

Charakteristikum	Studie 301
Population	<p>Ausgewählte Einschlusskriterien</p> <ul style="list-style-type: none"> • LCA-Diagnose aufgrund von Mutationen im Gen <i>RPE65</i>; Durchführung oder Bestätigung der Molekulardiagnostik erfolgte durch ein CLIA (Clinical Laboratory Improvement Amendments of 1988)-zertifiziertes Labor. • Alter ≥ 3 Jahre. • Sehschärfe schlechter als 20/60 (beide Augen) und/oder ein Gesichtsfeld kleiner als 20° in jedem Meridian (beide Augen), gemessen mit beispielsweise III4e Isopter oder äquivalent. • Ausreichend lebensfähige Netzhautzellen, bestimmt durch nicht-invasive Methoden wie einer optischen Kohärenztomographie (OCT) und/oder Ophthalmoskopie. Folgendes muss erfüllt sein: <ul style="list-style-type: none"> ○ Vorhandensein eines Bereiches der Netzhaut innerhalb des hinteren Pols von $> 100 \mu\text{m}$ Dicke sichtbar durch OCT oder ○ ≥ 3 Pupillendurchmesser der Netzhaut ohne Atrophie oder Pigmentdegeneration innerhalb des hinteren Pols oder ○ ein verbleibendes Gesichtsfeld innerhalb von 30° der Fixation, gemessen mit III4e Isopter oder gleichwertigem. • Der MLMT muss auswertbar sein (primärer Endpunkt): <ul style="list-style-type: none"> ○ Die Fähigkeit, den MLMT in den definierten Beleuchtungsstärken zu absolvieren. Personen mussten einen Genauigkeitswert¹⁾ von ≤ 1 bei einem Mobilitätstest zum Zeitpunkt des Screenings von 400 Lux oder weniger erreichen. Personen mit einem Genauigkeitswert¹⁾ von > 1 bei allen MLMT zum Zeitpunkt des Screenings oder die die Teilnahme verweigern, wurden von der Studienteilnahme ausgeschlossen. ○ Die teilnehmenden Personen dürfen den MLMT zum Zeitpunkt des Screenings bei einer Beleuchtungsstärke von 1 Lux nicht bestehen, um für die Studie geeignet zu sein; Personen, die einen oder mehrere MLMT zum Zeitpunkt des Screenings bei 1 Lux bestehen, sind von der Studienteilnahme ausgeschlossen. <p>Ausschlusskriterien</p> <ul style="list-style-type: none"> • Die Person ist nicht in der Lage oder nicht gewillt die Anforderungen der Studie zu erfüllen, was auch den Erhalt von bilateralen, subretinalen Injektionen des Vektors beinhaltet. • Jede vorherige Teilnahme an einer Studie, bei der ein gentherapeutischer Vektor verabreicht wurde. • Einsatz von Retinoid-Verbindungen oder deren Vorläufersubstanzen, die möglicherweise eine Interaktion mit der biochemischen Aktivität des Enzyms RPE65 eingehen könnten; Eine Teilnahme an der Studie war möglich, sofern diese Verbindungen für 18 Monaten nicht mehr angewandt wurden. • Vorangegangene intraokulare Operation innerhalb von sechs Monaten. • Bekannte Empfindlichkeiten gegenüber Medikamenten, die in der perioperativen Phase verwendet werden sollen. • Vorerkrankungen des Auges oder erschwerende systemische Erkrankungen, die die geplante Operation nicht zulassen, oder die Interpretation der Studie beeinträchtigen würden. Zu den erschwerenden systemischen Erkrankungen gehören solche, bei denen die Krankheit an sich oder die Behandlung der Erkrankung die Funktion des Auges verändern kann. Beispiele sind Malignome, deren Behandlung die Funktion des zentralen Nervensystems beeinflussen kann (z. B.: Strahlentherapie der Augenhöhle; Leukämie mit Beteiligung des Zentralen Nervensystems / Sehnervs). Erkrankte Personen mit Diabetes oder einer Sichelzellanämie wurden ausgeschlossen, wenn sie eine fortgeschrittene Retinopathie (z. B. Makulaödem oder proliferative Veränderung) hatten. Erkrankte Personen mit einer Immunschwäche (erworben oder angeboren) wurden ebenfalls ausgeschlossen, da eine

Charakteristikum	Studie 301
	<p>Prädisposition für opportunistische Infektionen bestehen kann (wie beispielsweise eine CMV Retinitis).</p> <ul style="list-style-type: none"> • Erkrankte Personen, die nicht in der Lage sind, einen Mobilitätstest aus anderen Gründen als schlechtem Sehen durchzuführen, einschließlich körperlicher und geistiger Einschränkungen.
Intervention und Zahl der Patienten	<p>Es wurden 36 Personen gescreent, wovon 31 Personen randomisiert wurden.</p> <p><u>Interventionsgruppe (N = 21)</u> Nicht-simultane Injektionen von $1,5 \times 10^{11}$ Vg AAV2-hRPE65v2 (Voretigen Neparvovec) mit einem Gesamtvolumen von 300 μl in den subretinalen Bereichen jedes Auges. Die sequentielle subretinale Injektion in beide Augen sollte innerhalb von 18 Tagen erfolgen.</p> <p><u>Kontrollgruppe (N = 10)</u> Keine Verabreichung der Prüfmedikation für einen Zeitraum von mindestens einem Jahr nach Baseline. Anschließend Erhalt der Prüfmedikation möglich, sofern Ein- und Ausschlusskriterien weiterhin erfüllt waren.</p>
Ort und Zeitraum der Durchführung	<p>Ort 2 Studienzentren in den USA:</p> <ul style="list-style-type: none"> • Children's Hospital of Philadelphia (CHOP), N = 19 • University of Iowa (IOWA), N = 12 <p>Studienzeitraum</p> <ul style="list-style-type: none"> • Einschluss erste/r Patient/in: 15.11.2012. • Letzte Visite erste/r Patient/in: 06.04.2015. • Einziger Datenschnitt vom 16.07.2015 (Interimsanalysen waren nicht geplant). Zusätzlich wurde im Studienbericht angegeben, dass die Datenbank für die Visite Jahr 1B/C zum 24.08.2015 geschlossen wurde. • Finaler Studienbericht: 13.12.2016, inklusive 3 administrative Amendments nach Veröffentlichung des Studienberichts (letztes Amendment vom 09.08.2017). <p>Die Nachbeobachtungen für alle Studienteilnehmenden, die die Prüfmedikation erhalten haben, sollen gemäß Studienprotokoll in Studie LTFU/302 für bis zu 15 Jahre laufen.</p>
Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte gemäß Studienprotokoll	<p>Primärer Wirksamkeitsendpunkt <u>MLMT</u>: Inwieweit durch die Administration des Prüfpräparats die Fähigkeit zu navigieren (untersucht mittels MLMT) bei Erwachsenen und Kindern ab 3 Jahren mit einer RPE65-Mutation verbessert wird. Änderung im MLMT bilateral gemessen nach Jahr 1B bzw. Jahr 1C zwischen den Behandlungsgruppen.</p> <p>Sekundäre Wirksamkeitsendpunkte</p> <ul style="list-style-type: none"> • Vollfeld-Lichtempfindlichkeits-Schwellenwert-Test (FST): Veränderung in der Lichtempfindlichkeit (Durchschnitt beider Augen) für weißes Licht nach einem Jahr (Jahr 1B/C) im Vergleich zur Baseline • Monokularer Mobilitätstest²⁾: Veränderung im Mobilitätsscore (unilateral, Durchführung mit dem Auge das zuerst behandelt wurde) nach einem Jahr (Jahr 1B/C) im Vergleich zu Baseline. • Sehschärfetest: Veränderung der Sehschärfe anhand der ETDRS- oder HOTV-Sehtafel (Durchschnitt beider Augen) nach einem Jahr (Jahr 1B/C) im Vergleich zu Baseline <p>Weitere Endpunkte</p> <ul style="list-style-type: none"> • Seh- und Netzhautfunktionstest <ul style="list-style-type: none"> ○ Pupillenreflexstest ○ Gesichtsfeldtests (Perimetrie nach Goldmann und Humphrey)

Charakteristikum	Studie 301
	<ul style="list-style-type: none"> ○ Kontrastempfindlichkeit²⁾ • Orientierungs- und Mobilitätsmessungen • Patientenberichteter Endpunkt zur visuellen Lebensqualität <p>Sicherheitsendpunkte (gemäß SAP)</p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse • Begleitmedikation • Körperliche Untersuchung • Ophthalmologische Untersuchungen • Klinisches Labor • Hämatologische Untersuchungen • Immunologische Untersuchungen • Vektorablösung³⁾
Subgruppenanalysen	Keine vorgesehen und durchgeführt.

- ¹⁾ Genauigkeitswert des MLMT: Anzahl Verstöße gegen die Genauigkeitsregeln dividiert durch Gesamtzahl der Hindernisse.
- ²⁾ Laut SAP (Version 4.0 vom 19.08.2015) war die monokulare Veränderung im MLMT nach einem Jahr (Jahr 1B/C) ein sekundärer Wirksamkeitsendpunkt (Durchführung mit dem Auge, das zuerst behandelt wurde), während der Endpunkt Kontrastempfindlichkeit als zusätzlicher Wirksamkeitsendpunkt erfasst wurde. Im Studienprotokoll (Version 6.0 vom 20.08.2013) verhält es sich umgekehrt: MLMT (unilateral) wird als zusätzlicher Endpunkt erwähnt, während die Kontrastempfindlichkeit als sekundärer Endpunkt definiert ist. Die Änderung wurde nach Rücksprache mit der FDA im aktuellen SAP dokumentiert.
- ³⁾ Das Vorhandensein des Vektors in peripherem Blut oder Tränensammlung.

Abkürzungen: DSMB: Data and Safety Monitoring Board; ETDRS: Early Treatment of Diabetic Retinopathy Study; FDA: U.S. Food and Drug Administration; Jahr 1B: Jahr 1 nach Behandlung des zweiten Auges in der Interventionsgruppe; Jahr 1C: Jahr 1 nach Baseline der Kontrollgruppe; LCA: Lebersche kongenitale Anomalie; MLMT: Multi-Luminanz-Mobilitätstest; SAP: Statistischer Analyseplan; UE: Unerwünschte/s Ereignis/se; Vg: Vektorgenome.

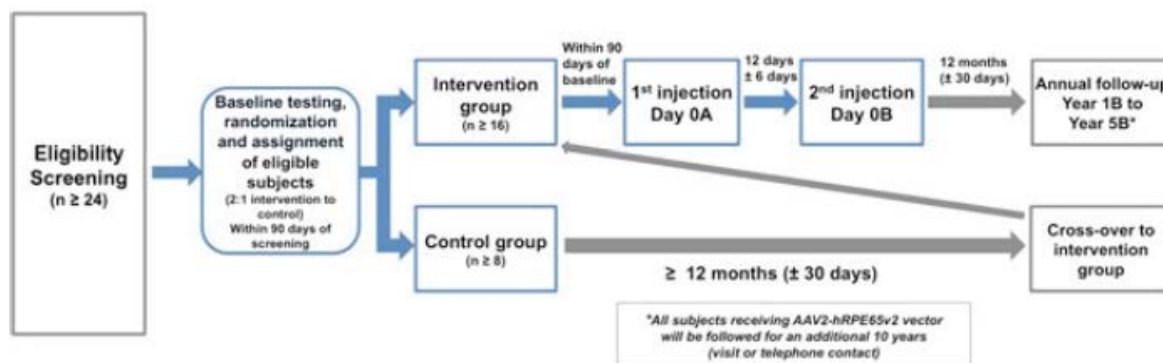


Abbildung 1: Studiendesign der Studie 301

Protokolländerungen zur Studie 301

Es gab für das Studienprotokoll 5 Änderungen zwischen Version 1 (19. Oktober 2009) und Version 6 (20. August 2013). Version 6 war die Protokollversion, die nach Einschluss des ersten Patienten / der ersten Patientin (15. November 2012) in Kraft getreten ist. Die Änderungen beinhalteten Klarstellungen bzw. administrative Änderungen. Wie viele Studienteilnehmende bei Inkrafttreten des Protokolls 6 bereits in die Studie eingeschlossen waren, ist nicht bekannt.

Tabelle 2: Charakterisierung der Intervention in der Studie 301

Intervention	Kontrolle
<p>Nicht-simultane Injektionen von je $1,5 \times 10^{11}$ Vg Voretigen Neparvovec (Luxturna®) mit einem Gesamtvolumen von 300 µl, in den subretinalen Bereich beider Augen.</p> <p>Der Vektor wurde in 1 ml Aliquots als Suspension in einer Konzentration von $5E12$ Vg/ml geliefert und enthält steriles Wasser mit 180 mM Natriumchlorid, 10 mM Natriumphosphat, 0,001 % Lutrol F68 bei einem pH-Wert von 7,3. Das Lösungsmittel wurde so zugesetzt, dass am Ende die zulassungsrelevante Konzentration von $1,5 \times 10^{11}$ Vg vorlag. Die Injektion in das zweite Auge sollte 12 ± 6 Tage nach der ersten Injektion erfolgen (Tag 0B).</p> <p>Die Vektorapplikation fand auf der ophthalmologischen chirurgischen Station des Studienzentrums unter Anästhesie statt. Ein/e qualifizierte/r Pharmazeut/in mit einem speziellen Training war für die Herstellung der Prüfmedikation zuständig.</p> <p>Der empfohlene Applikationsort lag im post-äquatoralem Bereich im hinteren Augenabschnitt. Es wurde $\frac{1}{3}$ bis $\frac{1}{4}$ der Gesamtfläche der Retina, inklusive eines Teils der Makula, die als Zielort des visuellen Sehens und der Empfindlichkeit angesehen wird, gesetzt. Die Prüfmedikation durfte nicht in die unmittelbare Umgebung der Fovea centralis verabreicht werden, um diese nicht zu beschädigen.</p> <p>Patientinnen und Patienten sollten 4–6 Stunden nach der Operation im Krankenhaus weiter beobachtet werden. Generell sollten sie nicht über Nacht im Krankenhaus bleiben, wurden aber gebeten das Bett nicht zu verlassen, mit Ausnahme zur Essensaufnahme und zum Aufsuchen des Bades. Sofern erforderlich, war ein längerer Krankenhausaufenthalt möglich.</p>	<p>Beobachtendes Abwarten für mindestens ein Jahr nach Baseline.</p>
<p>Nicht erlaubte Begleitmedikation</p> <ul style="list-style-type: none"> • Andere Studienwirkstoffe • Hochdosiertes Vitamin A (> 7.500 Retinol-Äquivalente oder > 3.300 IU am Tag) • Tretinoinhaltige Hautcreme (z. B. Retin-A®) • Isotretinoin • Viagra® (Sildenafil) oder ähnliche Substanzen zur Behandlung der erektilen Dysfunktion • Hydroxychloroquin, Chloroquin, Thioridazin oder ähnliche retinotoxische Substanzen <p>Erlaubte Begleitmedikation</p> <p>Studienteilnehmende im Interventionsarm erhielten systemische Glukokortikoide vor der 1. Vektorinjektion (Tag -3A) und kontinuierlich für die nächsten acht Tage bis zur 2. Injektion (Tag 8B). Je nach Zeitpunkt der zweiten Applikation konnten sich die Studienteilnehmenden zwischen 18 und 30 Tagen unter Glukokortikoid-Therapie befinden.</p> <p><u>Vorbehandlung</u></p> <ul style="list-style-type: none"> • Zur 1. Injektion von Voretigen Neparvovec: Prednison zur oralen Anwendung <ul style="list-style-type: none"> ○ 1 mg/kg/Tag (max. 40 mg/Tag) für 7 Tage ab dem 3. Tag vor der 1. Injektion ○ 0,5 mg/kg/Tag (max. 20 mg/Tag) für weitere 5 Tage ○ 0,5 mg/kg/QOD (max. 20 mg/Tag) bis 3 Tage vor der 2. Injektion • Zur 2. Injektion von Voretigen Neparvovec: Prednison zur oralen Anwendung <ul style="list-style-type: none"> ○ 1 mg/kg/Tag (max. 40 mg/Tag) für 7 Tage ab dem 3. Tag vor der 2. Injektion ○ 0,5 mg/kg/Tag (max. 20 mg/Tag) bis 8 Tage nach der 2. Injektion <p>Standardmedikation, die zu einer Operation verwendet werden, wie Anästhetika und Augentropfen (Antibiotika, Steroide und/oder Verdünnungstropfen oder Salben), wurden nicht im CRF dokumentiert, waren aber in den Krankenhausunterlagen vermerkt. Gleiches gilt für Augentropfen, die zur ophthalmologischen Untersuchung und für bestimmte Tests zur Überprüfung der retinalen Funktion bzw. Sehleistung eingesetzt wurden.</p>	

Abkürzungen: CRF: Case Report Form; QOD: alle zwei Tage; Vg: Vektorgenome.

2.5 Endpunkte und statistische Methoden

Laut SGB V § 35a Absatz 1 Satz 11 1. Halbs. gilt der medizinische Zusatznutzen durch die Zulassung als belegt. Nachweise gemäß § 5 Absatz 1 bis 6 sind nicht verpflichtend vorzulegen.

Im folgenden Kapitel werden schrittweise die Eignung der vorgelegten Endpunkte hinsichtlich Patientenrelevanz, Operationalisierung, Validität, Verzerrungspotential und statistischer Auswertung beurteilt. Dazu wurden das Nutzendossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen für die Endpunkte Perimetrie nach Humphrey und Goldmann, Vollfeld-Lichtempfindlichkeits-Schwellenwert-Test (FST), Sehschärfe mittels HOTV-Sehtafel durchgeführt.

2.5.1 Patientenrelevanz der Endpunkte

Tabelle 3: Bewertung der Patientenrelevanz der Endpunkte der Studie 301

Endpunkt	Kategorie	Durch den pU als patientenrelevant bewertet	In der Nutzenbewertung als patientenrelevant bewertet
Mortalität ¹⁾	Mortalität	Ja	Ja
Änderung im MLMT („Mobilitätstest“) ²⁾³⁾	Morbidität	Ja	Ja
Vollfeld-Lichtempfindlichkeits-Schwellenwert-Test („FST“)		Ja	Unklar
Perimetrie nach Humphrey		Ja	Unklar
Perimetrie nach Goldmann		Ja	Unklar
Sehschärfe anhand der ETDRS- oder HOTV-Sehtafel		Ja	Ja
Kontrastempfindlichkeit		Nein	Unklar
Orientierungs- und Mobilitätsmessungen		Nein	Ja
PRO-Instrument zur Erfassung der visuellen Lebensqualität ³⁾	Lebensqualität	Ja	Ja
Unerwünschte Ereignisse	Sicherheit	Ja	Ja ⁴⁾

¹⁾ Im Rahmen der Sicherheitserhebung erfasst.

²⁾ Primärer Endpunkt: Änderung im MLMT bilateral gemessen nach Jahr 1B (Jahr 1 nach Behandlung des zweiten Auges in der Interventionsgruppe) bzw. Jahr 1C (Jahr 1 nach Baseline der Kontrollgruppe) zwischen den Behandlungsgruppen.

³⁾ Vom pU entwickelter Fragebogen in Orientierung an den National Eye Institute Visual Functioning Questionnaire – 25 (NEI VFQ-25).

⁴⁾ Patientenrelevanz von Laborparametern unklar.

Abkürzungen: ETDRS: Early Treatment of Diabetic Retinopathy Study; FST: Vollfeld-Lichtempfindlichkeits-Schwellenwert-Test; MLMT: Multi-Luminanz-Mobilitätstest; PRO: Patient-Reported Outcome; pU: pharmazeutischer Unternehmer.

2.5.2 Charakterisierung der patientenrelevanten und ergänzend dargestellten Endpunkte

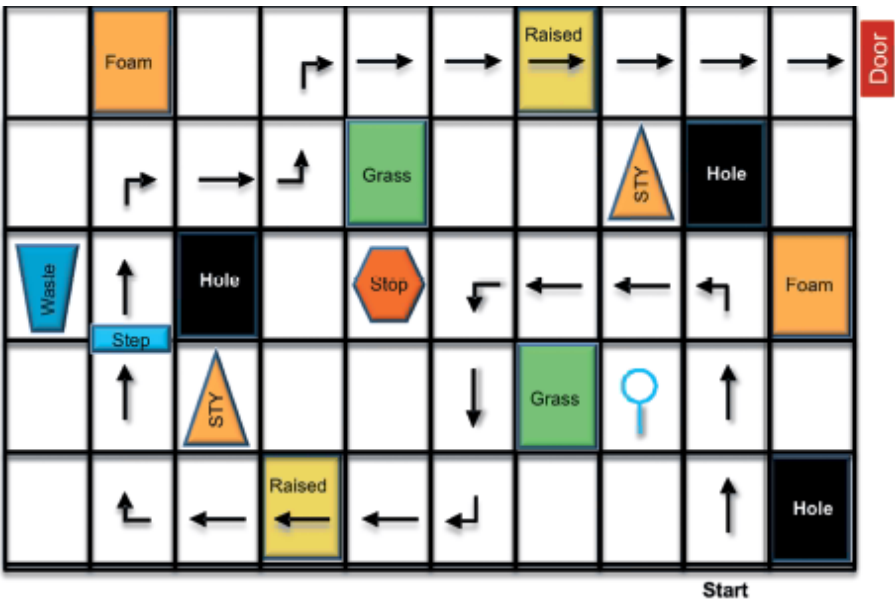
Im Folgenden werden die als patientenrelevant bzw. unklar patientenrelevant eingeschätzten Endpunkte (s. Tabelle 3) hinsichtlich ihrer Operationalisierung und Validität bewertet. Das Ergebnis der Betrachtung der Endpunkte findet sich aufgeteilt nach Endpunktkategorien in den Tabellen 4–8.

Tabelle 4: Beschreibung der Endpunkte der Kategorie Mortalität

Mortalität	
Operationalisierung	<p>Der pU gibt das Gesamtüberleben als Endpunkt der Kategorie Mortalität an. Der Endpunkt wurde als UE während der Studie erfasst und in der Interventionsgruppe ab Tag 0A (= erster Tag der 1. Injektion) bzw. ab dem ersten Tag der Baselineuntersuchungen in der Kontrollgruppe erhoben. Das Auftreten von UE wurde bei jeder Visite mittels Befragung oder Bericht des Patienten / der Patientin, körperlicher Beobachtung oder Untersuchung, oder mittels Labor- oder anderer diagnostischer Tests erfasst. Falls die Person vorab aus der Studie ausschied, wurde der Endpunkt als UE, sofern dieser behandlungsbedingt war, bis zu 30 Tage nachverfolgt.</p> <p>Auswertung/Auswertungspopulation Die Erhebung erfolgte auf Basis des finalen Datenschnitts (16.07.2015).</p> <p>Erhebungszeitpunkte Durchgehend von Tag 0A (erster Tag der Injektion bei der Interventionsgruppe) bzw. Baseline (erster Tag der Baseline-Untersuchung der Kontrollgruppe) bis zur Visite Jahr 1B/C.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nur zum Teil nachvollziehbar. Todesfälle wurden als UE erfasst. Kritisch anzumerken ist der kürzere Erfassungszeitraum der UE in der Interventionsgruppe, die nicht ab Baseline (wie in der Kontrollgruppe) sondern ab dem Tag der 1. Injektion gezählt wurden, wodurch Todesfälle in der Interventionsgruppe zwischen Screening und Injektion ins erste Auge in der Behandlungsgruppe (jeweils ≤ 90 Tage zwischen Screening und Baseline sowie Baseline und 1. Injektion) sowie zwischen Screening und Baseline in der Kontrollgruppe nicht erfasst werden konnten.</p> <p>Validität und Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der <u>Verlängerung des Überlebens</u>, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Die Erhebung der Mortalität als Wirksamkeitsparameter war kein Endpunkt in der Studie 301. Es ist jedoch darauf hinzuweisen, dass im Rahmen der Sicherheitsauswertung Todesfälle miterfasst wurden und somit – sofern zutreffend – dargestellt werden.</p>

Abkürzungen: AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; Jahr 1B: Jahr 1 nach Behandlung des zweiten Auges in der Interventionsgruppe; Jahr 1C: Jahr 1 nach Baseline der Kontrollgruppe; pU: pharmazeutischer Unternehmer; UE: Unerwünschte/s Ereignis/se.

Tabelle 5: Beschreibung der Endpunkte der Kategorie Morbidität

Morbidität	
Endpunkt	Multi-Luminanz-Mobilitätstest (MLMT)
Operationalisierung	<p>Der MLMT dient dazu Veränderungen im funktionalen Sehen zu messen, insbesondere die Fähigkeit sich unter verschiedenen Lichtverhältnissen in einem Hindernisparcours zu orientieren und selbstständig bewegen zu können.</p> <p><u>Aufbau</u> Die Gesamtfläche des Hindernisparcours beträgt 7 x 12 Fuß, wobei der Hindernisparcours auf einer 5 x 10 Fuß großen Fläche aufgestellt wird und er zusätzlich eine 1-Fuß große, an allen Ecken angrenzende, freie Fläche besitzt.</p> <p>Der Test verwendet sieben standardisierte Beleuchtungsstärken von 1 bis 400 Lux, die sich mit kalibrierten Lichtmessgeräten an fünf verschiedenen Positionen auf dem Hindernisparcours befinden. Für die Durchführung des Mobilitätstests wurden mithilfe von dreidimensionalen Hindernissen, mit hohem Kontrast zum Boden und zu den Stirnwänden, mit Farbunterschieden sowie mit Positionierung in unterschiedlichen Höhen, 12 Hindernisparcours mit vergleichbarem Schwierigkeitsgrad entwickelt (Abbildung 2).</p> <p>Für Erwachsene und ältere Kinder entsprach die Größe der Richtungspfeile der jeweiligen Sehschärfe, gemessen anhand einer Snellen-Sehtafel von 20/200 in einem Abstand von 2 m. Für jüngere Kinder (3–6 Jahre) konnten Allen-Card-Figuren eingesetzt werden, deren Größe ebenfalls der Sehschärfe von 20/200, in einem Abstand von 1–1,5 m aufgestellt, entsprach.</p>
	 <p>Abbildung 2: Beispiel für einen der 12 Hindernisparcours des MLMT [10]</p> <p><u>Durchführung des Tests</u> Vor den Baseline-/Studien-Visiten wurden Trainingsdurchläufe in allen drei Testsituation (bilateral und unilateral mit jedem Auge einzeln) bei einer Beleuchtungsstärke von 250 Lux (und ggf. niedriger) mit einem hierfür extra eingesetzten Übungsparcours durchgeführt. Sehhilfen konnten, sofern notwendig, bei den Übungsläufen eingesetzt werden. Die Teilnehmenden erhielten wesentliche Instruktionen und Anweisungen zur Durchführung des Tests.</p> <p>Bei der eigentlichen Durchführung des Tests während der Studie sollten sich die Teilnehmenden in einer 40-minütigen Phase bei niedriger Beleuchtung an die Dunkelheit gewöhnen. Danach wurde zunächst mit dem ersten, dann mit dem zweiten Auge und zuletzt mit beiden Augen unter mindestens zwei unterschiedlichen Lichtverhältnissen</p>

Morbidität

ein Hindernisparcours durchlaufen. Sofern es eine starke Abweichung im Testergebnis zwischen beiden Augen gab, konnte ein weiterer Testdurchlauf stattfinden. Die Tests sollten alle am gleichen Tag stattfinden, allerdings bestand die Möglichkeit Durchläufe am Folgetag durchzuführen. Es gab vor der Visite eine randomisierte Auswahl an Hindernisparcours für jede Augen- und Lichtkombination. Falls eine Person vom Kurs abkam und sich nach zwei Schritten nicht selbst korrigieren konnte oder durch eine Kollision gefährdet war, führte die Testperson die Person zurück auf den Kurs. Es wurden keine weiteren verbalen oder physischen Hinweise gegeben. Gemäß Dossier M4 waren keine Hilfsmittel (Gehstock, Sehhilfe etc.) während der Durchführung des Tests für beide Behandlungsgruppen erlaubt. Jeder Testdurchlauf wurde von zwei Personen (bei Dissens drei Personen) auf Basis von Videoaufnahmen bewertet. Im Manual sind zusätzlich Vorgaben zur Verwendung der Kamera angegeben. Gemäß pU wurde das Ergebnis des Tests verblindet bewertet. Informationen zur Gruppenzuteilung oder zu anderen Ergebnissen von retinalen/visuellen Funktionstests lagen den bewertenden Personen nicht vor. Zur standardisierten Bewertung der Tests fanden Schulungen statt.

Genauigkeitswert und Zeitwert

Um den Durchlauf bei einer Beleuchtungsstärke zu bestehen, mussten sowohl die Genauigkeits- als auch die Zeitkriterien bestanden werden. Dies trifft auf alle Test-situationen (verdeckte/unverdeckte Augen) zu. Der Genauigkeitswert wurde berechnet, indem die Anzahl der Verstöße gegen die Genauigkeitsregeln durch die Gesamtzahl der Hindernisse dividiert wurde. Um laut Genauigkeitswert zu bestehen, musste eine Genauigkeitsbewertung $\leq 0,25$ erreicht werden. Das bedeutet, dass maximal drei Fehler begangen werden durften.

Der Zeitwert wurde durch die Kombination der Sekunden zur Fertigstellung des Parcours mit erteilten Zeitstrafen ermittelt. Die Zeitstrafen wurden nach einem im Voraus festgelegten Schema vergeben (z. B. 15 Sekunden für das Verlassen der Strecke). Zeitstrafen wurden zur tatsächlich benötigten Zeit für einen Durchlauf addiert. Um einen Durchlauf auf Basis dieses ermittelten Zeitwerts zu bestehen, musste dieser Wert unter 180 s liegen.

Auswahl der Beleuchtungsstärken

Die unterschiedlichen Beleuchtungsstärken sind laut Studienunterlagen alltäglichen Lichtverhältnissen nachempfunden (Abbildung 3).

Während des Studienverlaufs wurden zwei Grenzwerte zu jeder Visite ermittelt:

- Niedrigster Lichtsensitivitätsgrenzwert: Beleuchtungsstärke, bei der die Person den Test nur knapp besteht.
- Sub-Sensitivitätsgrenzwert Beleuchtungsstärke, die unter dem niedrigsten Sensitivitätsgrenzwert liegt. Dies ist die Beleuchtungsstärke, bei der die Person nicht mehr selbstständig navigieren kann.

Lux	Leuchtdichte (cd/m ²)	Entsprechende Umgebung
1	0,32 Mesopisches Sehen	Mondlose Sommernacht; Innenraum-Nachtbeleuchtung
4	1,3 Mesopisches Sehen	Wolkenlose Sommernacht mit Halbmond; Parkplatz bei Nacht
10	3,2 Mesopisches Sehen	60 min nach Sonnenuntergang in städtischer Umgebung; Bushaltestelle bei Nacht
50	15,9 Photopisches Sehen	Außenbereich eines Bahnhofs bei Nacht; schwach beleuchtetes Treppenhaus
125	39,8 Photopisches Sehen	30 min vor einem wolkenlosen Sonnenaufgang; Innenraum eines Einkaufszentrums, Zuges oder Busses bei Nacht
250	79,6 Photopisches Sehen	Innenraum eines Aufzuges, Büroflurs oder einer Bibliothek
400	127,3 Photopisches Sehen	Büroräume

Abbildung 3: Verwendete Beleuchtungsstärken beim MLMT der Studie 301

Morbidität

Um eine Beleuchtungsstärke zu Baseline zu definieren, wurde das Licht bei dem Hindernisparcours für alle drei Testsituationen (mit beiden Augen, jeweils mit einem verdeckten Auge) solange gedimmt, bis die Person der Meinung war, den Test kaum durchführen zu können. Die Person konnte sich währenddessen im Raum bzw. auf dem Hindernisparcours bewegen. Anschließend wurde der Test bei dieser Beleuchtungsstärke durchgeführt. Falls die Person den Test bestand, wurde der Test bei der nächstniedrigeren Beleuchtungsstärke durchgeführt, bis die Person den Test in der jeweiligen Testsituation nicht mehr bestehen konnte.

Bei den darauffolgenden Visiten wurde in Abhängigkeit der vorhergehenden Visite die Beleuchtungsstärke gewählt. Es waren mindestens 2 Durchläufe pro Testsituation (erstes Auge, zweites Auge, bilateral) erforderlich: ein Durchlauf beim niedrigsten Sensitivitätsgrenzwert der vorherigen Visite und ein Durchlauf bei einer neuen Beleuchtungsstärke. Das Layout der Hindernisparcours wurde für jeden Durchlauf geändert und die Kurskonfiguration wurde vor jeder Visite randomisiert.

Falls beispielsweise ein/e Patient/in zu Baseline den Test bei 125 Lux bestanden hat, wurde bei der folgenden Visite ebenfalls bei 125 Lux und – sofern bestanden – anschließend bei 50 Lux (der nächstkleineren Beleuchtungsstärke) getestet. Falls der Test bei 125 Lux bereits nicht bestanden wurde, wurde die nächst höhere Beleuchtungsstärke verwendet.

Maßgeblich für die Mobilität war die Beleuchtungsstärke, bei der der/die Patient/in den Mobilitätstest in der entsprechenden Testsituation mit verdecktem/unverdecktem Auge noch bestehen konnte. Dieser Lux-Wert wurde dann für jede Testsituation in einen Mobilitätsscore zwischen 6 und -1 umgewandelt, wobei ein höherer Score einer besseren Mobilität entspricht (Tabelle 6).

Tabelle 6: Umrechnungstabelle für den Mobilitätsscore

Lux	1	4	10	50	125	250	400	> 400
Mobilitätsscore	6	5	4	3	2	1	0	-1

Qualitätssicherung der Bewertung der Tests während der Studie

Während der Studie war geplant sowohl die Interrater- und die Intrarater-Reliabilität als auch die Konsentierung zwischen den Bewertungspersonen zu untersuchen, indem sowohl für die Kriterien des Genauigkeitswertes (wie beispielsweise Anzahl an nicht erkannten Hindernissen, Häufigkeit der Korrekturen), des Zeitwertes als auch des finalen Bewertungskriteriums „bestanden / nicht bestanden“ Intra-Class-Korrelationen, Kappa-Werte sowie Anzahl an Übereinstimmungen berechnet wurde. Alle 3 Monate wurden 10 % der Videos vom vorherigen Quartal zufällig für eine erneute Beurteilung ausgewählt und unter die neu zu bewertenden Videos gemischt, ohne dass es die Bewertungspersonen wussten.

Auswertung/Auswertungspopulation

Die Erhebung erfolgte anhand der ITT-Population für den Datenschnitt 16.07.2015.

Als primärer Endpunkt wurde die Änderung im MLMT-Score bilateral (für beide unverdeckten Augen) zu Jahr 1B/C im Vergleich zu Baseline zwischen den Behandlungsgruppen untersucht.

Für den sekundären Endpunkt wurde das unilaterale (mit dem ersten Auge) Testszenario verwendet. Zusätzlich wurden post hoc auf Wunsch der FDA für die bilaterale Testsituation weitere Analysen durchgeführt: Zeit zum Durchlaufen des Parcours, Ergebnis in Abhängigkeit der 12 Parcours-Layouts sowie das Alter der Teilnehmenden. Zusätzlich wurde auch der Summenscore für alle drei Testbedingungen (erstes, zweites Auge und bilateral) untersucht.

MID

Als MID wurde durch den pU eine Verbesserung des Scores um ≥ 1 Punkte definiert.

Morbidität	
Erhebungszeitpunkte	<p>Zu Baseline, Monate 1, 3 und 6 sowie Jahr 1, wobei der primäre Endpunkt zum Studienende zu Jahr 1 erhoben wurde. Bei Studienabbruch waren keine weiteren Erhebungen vorgesehen.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nachvollziehbar. Positiv anzumerken ist die verblindete Beurteilung der Testdurchläufe durch 2 (bzw. optional 3) Bewertungspersonen bei einem offenen Studiendesign.</p> <p>Validität Der pU hat neben der Publikation von Chung et al. (2018) [3] den Bericht zur Validierungsstudie für die Nutzenbewertung eingereicht [23]. An der Validierungsstudie nahmen 60 Personen teil, von denen 31 eine Sehbeeinträchtigung aufwiesen und 29 normalsichtig waren. An allen vier Visiten (Visite 1: Screening, Visite 2 + 3 : Baseline-Untersuchungen, Visite 4: Jahr 1 nach Baseline) nahmen 53 (88 %) der Studienteilnehmenden teil (26 mit normaler Sicht, 27 mit eingeschränkter Sicht). Bei 20 Personen wurde eine LCA und bei 4 eine Retinitis pigmentosa diagnostiziert. Bei einer Person wurde eine RPE65-Mutation festgestellt. Es wurden Personen im Alter zwischen 4 und 40 Jahren eingeschlossen. Das Durchschnittsalter betrug 15,4 Jahre (SD: 9,4). In der Studie war geplant die Reliabilität, Validität sowie die Änderungssensitivität zu untersuchen. [3,23]</p> <p><u>Durchführung der Validierungsstudie</u> Die Durchführung der Validierungsstudie ähnelte der Durchführung und Bewertung während der Studie 301. Ziel war es, den Sub-Sensitivitätsgrenzwert und den niedrigsten Lichtsensitivitätsgrenzwert zu bestimmen. Es wurden insgesamt 12 zufällig ausgewählte Durchläufe für jede Visite ausgesucht. Dadurch konnte der/die Studienteilnehmende nicht mehr als 12 Durchläufe pro Visite durchführen (d. h. 4 Beleuchtungsstärken für jede der drei Augenszenarien). Auf jeder Lichtebeleg wurden die Teilnehmenden mit jedem Auge einzeln und mit beiden Augen unverdeckt getestet. [3,23] Die Beleuchtungsstärken wurden in ordinalskalierte Werte umgewandelt (Tabelle 6). In der vom pU mitgelieferten Validierungsstudie wurde für die Beleuchtungsstärken 100 und 150 Lux beim Mobilitätsscore der gleiche Wert (Mobilitätsscore: 2; Beleuchtungsstärke: 125 Lux) kodiert sowie für die Beleuchtungsstärken 200 und 250 Lux der Mobilitätsscore von 1 bzw. die Beleuchtungsstärke 250 Lux verwendet, um die Bedingungen der Phase-III-Studie nachzuahmen [23].</p> <p><u>Ergebnisse der Validierungsstudie</u> Zur Entwicklung des Tests wurden keine Daten vorgelegt. Aus den Protokollversionen der Studie 301 geht hervor, dass die Beleuchtungsstärken in Abhängigkeit der Test-Retest-Reliabilität ausgewählt wurden. Alle normalsichtigen Personen (n = 29) konnten den Test bei 1 Lux, der niedrigsten Beleuchtungsstufe, bestehen. Die Anzahl an verwendeten Beleuchtungsstärken bei sehbeeinträchtigten Personen lässt auf eine schiefe Verteilung schließen: Personen wurden überwiegend bei 1 Lux (n = 22) bzw. bei 4 und 10 Lux (jeweils n = 13) getestet und nur wenige Personen bei 50 Lux (n = 9), 100 Lux (n = 6), 150 Lux (n = 3) und bei 400 Lux (n = 3), keine Person bei 200 bzw. 250 Lux. Mehr als die Hälfte der Studienteilnehmenden haben den Test bei niedrigen Beleuchtungsstufen (bzw. hohen Mobilitätsscores) wiederholt, was auf Deckeneffekte des Tests hindeutet [23]. Da der Großteil der Teilnehmenden sich auf die niedrigen Beleuchtungsstufen verteilte, konnten die Beleuchtungsstufen 200 und 250 Lux nicht eingesetzt werden. Es kann für die Beleuchtungsstärken keine abschließende Validität und Reliabilitätsbeurteilung durchgeführt werden. In der Publikation von Chung et al. [3] wurde als eine der Einschränkungen des Fragebogens angegeben, dass es keine logarithmische Abstufung in den Leuchtstufen gab. Die Abstände orientierten sich stattdessen an Beleuchtungsstärken aus dem Alltag.</p>

Morbidität	
	<p>Für die Bewertung der Interrater-/Intrarater-Reliabilität wurden alle Videos (n = 3.168), bei der der MLMT eingesetzt wurde, bis zum 23.01.2015 inklusive der Validierungsstudie und der Studie 301 bewertet [23]. Im Ergebnis zeigte sich für alle untersuchten Komponenten des Mobilitätstests und für verschiedene Altersgruppen über die zwölf einzelnen Parcours-Layouts hinweg ähnlich hohe Werte für die Interrater- bzw. Intrarater-Reliabilität (z. B. finales Kriterium „bestanden / nicht bestanden“: kappa > 0,9) [3,23]. Zusätzlich wurde für die Validierungsstudie bei der bilateralen Testsituation für das finale Kriterium „bestanden / nicht bestanden“ eine Übereinstimmung von 88 % festgestellt [3]. Eine separate Auswertung der Reliabilität, unabhängig von der Studienpopulation aus der Phase-III-Studie, konnte nur zum Teil für die Validierungsstudie identifiziert werden, weshalb die Beurteilung der Interrater-/Intrarater-Reliabilität nicht abschließend möglich ist.</p> <p>In der Validierungsstudie konnte zwischen beiden Baseline-Werten bei 7 der 31 Personen mit Sehbeeinträchtigungen eine Verbesserung beobachtet werden, die aber zu Jahr 1 nicht anhielt. Dies deutet einerseits auf einen Lerneffekt bei kurz aufeinander folgenden Visiten, aber auch auf eine geringe Messunsicherheit für die gesamte Studie hin. Sowohl das Alter als auch die 12 Parcours-Layouts zeigten keinen Einfluss auf die Bewertungskriterien des Tests (u. a. Genauigkeits-, Zeitwert) [23].</p> <p>Der Test unterscheidet zwischen normalsichtigen und sehingeschränkten Personen. So zeigte sich bei den sehbeeinträchtigten Personen ein breites Spektrum an korrekten und fehlerhaften Kursdurchläufen, während sich im Gegensatz dazu bei normalsichtigen Personen keine Unterschiede zeigten. [3,23]</p> <p>Zur Beurteilung der Konstruktvalidität wurden die Genauigkeitswerte der Teilnehmenden in den Teststrecken mit den Testergebnissen zur Sehschärfe (anhand logMAR-Tafeln) verglichen. Dabei wurden auch Korrelationen für die visuell beeinträchtigten Personen gebildet, die moderat bis stark ausgeprägt waren (0,75–0,86), während sowohl für die Goldmann- als auch für die Humphrey-Perimetrie nur schwache bis moderate Korrelationen berechnet werden konnten (r = 0,37–0,64) [3]. Es wurde zusätzlich ein vom pU entwickelter Fragebogen zur visuellen Lebensqualität untersucht. Da die Validität des Fragebogens unklar ist, ist die Aussagesicherheit des Fragebogens unklar, weshalb der Fragebogen zur Untersuchung der Konstruktvalidität des MLMT als kritisch angesehen wird [3].</p> <p>Änderungssensitivität und MID</p> <p>Es wurde darüber hinaus die Fähigkeit, Veränderungen zu erkennen, untersucht. Unter bilateralen Testbedingungen konnte bei allen 26 normalsichtigen Personen nach einem Jahr eine Veränderung im MLMT-Score von 0 Punkten festgestellt werden, was ebenfalls bei 20 von 28 (71 %) der sehbeeinträchtigten Patienten zutraf. Bei den sehbeeinträchtigten Personen mit der Diagnose LCA bzw. Retinitis pigmentosa (n = 8) zeigte sich unter bilateralen Testbedingungen eine Verschlechterung von 1 oder 2 Punkten. Bei keiner der Personen konnte nach einem Jahr eine Verbesserung beobachtet werden [3,23].</p> <p>Der pU definiert eine absolute Änderung um 1 Punkt im Mobilitätsscore gemäß Umrechnungsskala (vgl. Tabelle 6) als eine klinisch relevante Änderung, ohne diese in Relation zu Veränderungen mit anderen validierten Instrumenten zu setzen. Da keine geeigneteren Verfahren, wie beispielsweise ankerbasierte Verfahren, zur Herleitung der MID herangezogen wurden, wird auf eine Darstellung der Ergebnisse anhand der vom pU vorgegebenen MID verzichtet [23].</p> <p>Patientenrelevanz</p> <p>Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p>

Morbidität	
	<p>Die Orientierung und Mobilität unter unterschiedlichen Lichtverhältnissen wird als patientenrelevant angesehen, da diese direkt wahrnehmbar ist. Dabei wird die Patientenrelevanz der bilateralen Testsituation gegenüber der anderer Testszenarien mit verdeckten Augen als übergeordnet bewertet, da diese der Alltagssituation von betroffenen Personen am nächsten ist. Da sich die vom pU zusätzlich durchgeführten Post-hoc-Analysen zu den Bewertungskriterien des Tests in der Gesamtbewertung widerspiegeln, werden diese nicht dargestellt.</p> <p>Zusammenfassung</p> <p>Die Testdurchführung wurde anhand der SOP umfassend beschrieben und lässt auf ein standardisiertes Vorgehen schließen, dass auch die Teilnahme von Kindern berücksichtigt. Auch ist die verblindete Bewertung bei offenem Studiendesign zu begrüßen. Zudem wurden Maßnahmen zur Qualitätssicherung während der Studie bei der verblindeten Bewertung der Testdurchläufe und eine Randomisierung der Hindernisparcours zur Reduktion von Lerneffekten ergriffen. In der Validierungsstudie deuten die Ergebnisse auf eine geringe Messunsicherheit und gute Reliabilität hin. Es konnte zudem eine ausreichende Konstruktvalidität nachgewiesen werden. Der Test konnte zwischen gesunden Personen und Personen mit Seheinschränkungen unterscheiden.</p> <p>Aus diesen Gründen werden die Ergebnisse trotz der unten stehenden kritischen Anmerkungen des bilateral gemessenen MLMT in der Nutzenbewertung dargestellt.</p> <p>Konkrete Angaben zur Entwicklung des Testparcours und der Untersuchung der Validierungsparameter zum Mobilitätsscore und zur finalen Bewertung „bestanden / nicht bestanden“ bei jeder Beleuchtungsstärke konnten nicht identifiziert werden. Auch kann eine abschließende Beurteilung des Tests bei den beiden Beleuchtungsstärken 200 und 250 Lux, die später zu 250 Lux zusammengefasst wurden, nicht vorgenommen werden. Die hohen Anzahl an durchgeführten Tests bei niedrigen Beleuchtungsstärken bzw. hohen Mobilitätsscores zeigt zudem Deckeneffekte an. Auch eine separate Darstellung der Interrater-/Intrater-Reliabilität der verwendeten Studien (u. a. Validierungsstudie, Studie 301) konnte nur zum Teil identifiziert werden. Es liegen zudem lediglich Hinweise auf eine vorhandene Änderungssensitivität des Instruments unter bilateralen Bedingungen vor. Eine validierte klinische Relevanzschwelle ist nicht vorhanden.</p>
Endpunkt	Vollfeld-Lichtempfindlichkeits-Schwellenwert-Test (FST)
Operationalisierung	<p>Der FST (steht für Full-field Light Sensitivity Threshold Test) hat das Ziel die subjektive Lichtempfindlichkeit des gesamten Gesichtsfeldes, bei der die Testperson noch sehen kann, zu erfassen.</p> <p>Angaben zur Durchführung des Tests wurden vereinzelt erst im Studienbericht bzw. im Dossier (Modul 4), nicht aber im Studienprotokoll, gemacht. Ein Handbuch, Manual bzw. eine SOP wurden ebenfalls nicht eingereicht.</p> <p>Es wurde ein Ganzfeld-Elektroretinogramm (ERG) angewandt, mit dem die Leuchtdichte eines Lichtblitzes, den die Testperson noch sehen kann, erfasst wird. Die Augen wurden zunächst erweitert. Der Test startete nach einer 40-minütigen Eingewöhnungsphase an die Dunkelheit. Es wurde jedes Auge einzeln getestet, indem das kontralaterale Auge verdeckt wurde. Zu Beginn der Prüfung blinkte eine Lampe in der gesamten Kuppel, begleitet von einem Piepton. Jedes Mal, wenn ein Piepton ertönte, musste der Teilnehmende durch Drücken einer "Ja"- oder "Nein"-Taste angeben, ob das Licht gesehen wurde oder nicht. Die Lichtblitze wurden mit unterschiedlichen Intensitäten (bis zu 80 dB Differenz) fortgesetzt. Ein Algorithmus identifizierte die minimale Leuchtdichte (Helligkeit), bei der die Testperson zuverlässig Licht wahrnahm. Die Leuchtdichte wurde in einen logarithmischen Wert umgewandelt. Für $\log_{10}(\text{cd s/m}^2)$ entspricht ein negativeres Ergebnis einem niedrigeren Schwellenwert und damit einer verbesserten Lichtempfindlichkeit, was auf eine verbesserte Photorezeptorfunktion hinweist. In der Phase-III-Studie wurden weiße, rote und blaue Lichtreize für jedes Auge einzeln getestet. Die Reihenfolge der Lichtintensität wurde durch Zufallsprinzip festgelegt.</p>

Morbidität	
	<p>Der pU stellt im Dossier als patientenrelevanten Endpunkt nur den sekundären Endpunkt dar, gemessen mit weißem Licht gemittelt auf beide Augen. Zusätzliche Analysen mit rotem und blauem Licht wurden ebenfalls im Studienbericht dargestellt.</p> <p>Subjektive Tests konnten zu Baseline mehrmals am Tag bzw. auf mehrere Tage verteilt gemessen werden, bis akzeptable Reliabilitätswerte erreicht wurden. Bei Kindern sollte die Dauer der Testdurchführungen kürzer sein, dafür an mehreren Tagen stattfinden. Unklar bleibt, ob ein ähnliches Vorgehen zu den weiteren Studienvisiten vorgesehen war und welche Reliabilitätskriterien verwendet wurden.</p> <p>Auswertung/Auswertungspopulation Die Auswertung erfolgte auf Basis der ITT-Population für den Datenschnitt 16.01.2015.</p> <p>Erhebungszeitpunkte Zu Baseline, Monate 1, 3, und 6 sowie Jahr 1. Eine Nachverfolgung im Rahmen der Studie 301 war nicht geplant, die Studienteilnehmenden konnten an der Langzeitbeobachtungsstudie LTFU/302 teilnehmen.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung kann nicht nachvollzogen werden. Unklar ist, warum sich erste Angaben zur Durchführung erst in Studienbericht und Herstellermodul M4 finden. Da es sich um eine unverblindete Studie handelt und laut Dossier auch keine verblindete Auswertung vorgesehen war, sind Angaben zur konkreten Testsituation im Studienprotokoll notwendig. Dies trifft insbesondere auf multizentrische Studien sowie auf Studien mit Teilnahme von Kindern zu, da eine erfolgreiche Durchführung des Tests sowohl von der Motivation als auch vom Verstehen der Testsituation und der Handlungsanweisung abhängen kann. Angaben, inwieweit das Studienpersonal in der Durchführung des Tests geschult war, konnten ebenfalls nicht identifiziert werden. Ähnlich wie beim MLMT erfolgt eine Adaption der an der Studie teilnehmenden Person an die Dunkelheit. Unklar bleibt, ob eine Reihenfolge bei der Erhebung der primären, sekundären und zusätzlichen Endpunkte vorgesehen war und ob mehrere Wiederholungen des Tests, ähnlich wie zu Baseline, auch zu anderen Visiten möglich war.</p> <p>Validität Nach Angaben des pU ist der Test geeignet zur Untersuchung der Lichtempfindlichkeit bei Personen mit Nystagmus, kleinem Gesichtsfeld und schlechter Sehschärfe. Vom pU wurde eine Studie von Roman et al. (2015) zur Entwicklung des Tests zur Quantifizierung des Sehvermögens bei Menschen mit Netzhautdegeneration, die zwischen 9 und 81 Jahre alt waren, vorgelegt [20]. In dieser Studie wurden 146 Personen mit Augenerkrankungen (u. a. Retinitis pigmentosa (n = 78), LCA oder früh einsetzende Netzhautdegeneration (n = 15)) einem FST für jeweils ein Auge unterzogen. Es konnte eine hohe Korrelation (r = 0,96) mit der kinetischen Perimetrie nach Goldmann sowie eine Wiederholbarkeit der Ergebnisse festgestellt werden [20]. In einer orientierenden Recherche konnte eine Studie von Messias et al. (2013) [15] identifiziert werden, die den Test u. a. mit einer statischen Perimetrie zur Gesichtsfeldmessung vergleicht sowie die Sehschärfe mit ETDRS-Sehtafel untersucht. Es nahmen 47 Personen im Alter zwischen 18 und 60 Jahren mit Retinitis pigmentosa teil. Eine moderate Korrelation zwischen statischer Perimetrie und FST (r = -0,39) sowie zwischen FST und Sehschärfe (r = 0,42; p < 0,001) konnte gezeigt werden. Es zeigte sich auch eine moderate Korrelation zwischen Testdauer und FST (r = -0,575; p < 0,001), was gemäß den Autoren darauf hindeutet, dass der Erfolg von der Länge der Messung abhängt [15]. Insgesamt konnten Reliabilitätskriterien erfüllt werden, wenngleich nur eine ausreichende Konstruktvalidität nachgewiesen werden konnte, wobei nur erwachsene Personen eingeschlossen wurden. Untersuchungen bei Kindern wurden nicht durchgeführt, wodurch eine abschließende Beurteilung der psychomerischen Eigenschaften in dieser Patientenpopulation derzeit nicht möglich ist.</p>

Morbidität	
	<p>Sowohl Informationen zur Änderungssensitivität als auch zu Reliabilitätskriterien des Instruments in der Studie 301 konnten nicht identifiziert.</p> <p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p> <p>Die Lichtempfindlichkeit ist ein patientenrelevanter Endpunkt. Die Patientenrelevanz ist aufgrund der künstlichen Messsituation unklar.</p> <p>Zusammenfassung Der Endpunkt wird in der Nutzenbewertung nicht berücksichtigt. In der Gesamtschau sind die Informationen zur Operationalisierung nicht ausreichend und die Durchführung/Standardisierung des Tests ist nicht nachvollziehbar. In Anbetracht des offenen Studiendesigns und der Teilnahme von Kindern ist eine standardisierte Durchführung essentiell, um valide und unverzerrte Messungen durchzuführen.</p> <p>Angaben zu Reliabilität, Änderungssensitivität und Validität insbesondere bei Einsatz des Instruments bei Kindern konnten ebenfalls nicht identifiziert werden.</p>
Endpunkt	
	Gesichtsfeldmessung mittels Perimetrie
Operationalisierung	
	<p>Zur Messung des Gesichtsfeldes wurden statische (nach Humphrey) und kinetische (nach Goldmann) Untersuchungsmethoden eingesetzt. Die Goldman-Perimetrie erfasst das gesamte Gesichtsfeld, während die Humphrey-Perimetrie sich auf bestimmte Regionen im Gesichtsfeld fokussiert. Da die Prüfmedikation in eine bestimmte Region (Makula) des Auges appliziert wird, werden bei der Humphrey-Perimetrie Unterschiede in der Funktion in dieser Region vor und nach Applikation der Prüfmedikation untersucht. Da eine Applikation in der Fovea-Region vermieden werden soll, wurde ebenfalls eine Untersuchung in dieser Region vorgenommen.</p> <p>Subjektive Tests konnten zu Baseline mehrmals am Tag bzw. auf mehrere Tage verteilt gemessen werden bis akzeptable Reliabilitätswerte erreicht wurden. Bei Kindern sollte die Dauer der Testdurchführungen kürzer sein, dafür an mehreren Tagen stattfinden. Unklar bleibt, ob ein ähnliches Vorgehen zu den weiteren Studienvisiten vorgesehen war und welche Reliabilitätskriterien verwendet wurden.</p> <p><u>Perimetrie nach Goldmann</u> Bei den kinetischen Untersuchungsmethoden nach Goldmann wird ein Ziellicht als visueller Reiz auf die Innenfläche einer Halbkugel projiziert. Die Intensität des Ziellichts bleibt konstant, das Ziellicht ist jedoch beweglich. Das Ziellicht wird von außerhalb der Gesichtsfeldgrenze in das vermutete Gesichtsfeld hineinbewegt und der Ort der erstmaligen Wahrnehmung wird dokumentiert.</p> <p>Zur Durchführung des Tests wird auf eine Publikation von Talib et al (2018) hingewiesen, die sich mit der Durchführung der Perimetrie nach Goldmann bei genetischen Therapien befasst [27]. Für die Durchführung wird eine hohle Halbkugel mit einem Radius von 30 cm und einer fixierten Hintergrundbeleuchtung verwendet. Die Testperson platziert sich davor. Auf die Innenseite der hohlen Halbkugel wird ein Licht erzeugt, das auf die Innenseite der hohlen Halbkugel als visueller Stimulus projiziert wird und sich von außen zur Mitte der Halbkugel bewegt. Sowohl die Hintergrundbeleuchtung als auch die Größe und Lichtintensität des Stimulus kann variiert werden. [27]. Sobald der Stimulus erscheint bzw. verschwindet und wieder erscheint soll die Testperson einen Knopf drücken. Dadurch kann das Ansprechen auf einen speziellen visuellen Stimulus genutzt werden, um die Grenzen des Gesichtsfeldes für diesen Stimulus festzulegen.</p> <p>Als Stimuli wurde laut Herstellermodul M4 in der Studie 301 der Stimuli V4e (Größe: 64 mm², Leuchtdichte: 315 cd/m²) und III4e (Größe: 4 mm², Leuchtdichte: 315 cd/m²) verwendet. Laut Herstellermodul M4 wurden die auf der Halbkugel gewonnenen Mess-</p>

Morbidität	
	<p>werte in Form eines Isopters (Linie der optischen Wahrnehmung von Punkten bei gleicher visueller Empfindlichkeit) auf eine ebene Fläche projiziert. Diese ebene Fläche bildet graphisch das Gesichtsfeld ab und ist in 24 Meridiane eingeteilt. Für jeden Meridian wird die Gradzahl der ermittelten Grenzen des Gesichtsfeldes bestimmt und schließlich die Gesamtsumme ermittelt.</p> <p>In der vom pU zitierten Publikation von Talib et al. (2018) [27] wird angegeben, dass eine geschulte Person den Test durchführen und bewerten soll, hierzu konnten für die Studie 301 keine Angaben identifiziert werden. Zudem konnten sowohl in den Studienunterlagen als auch im Herstellermodul M4 keine konkreten Angaben zur Durchführung des Tests (u. a. Geräteart, Standardisierung der Testsituation) identifiziert werden.</p> <p>Statische Perimetrie nach Humphrey Im Dossier wird zur Perimetrie nach Humphrey aufgeführt, dass der Aufbau des Humphrey-Perimeters dem des Goldmann-Perimeters entspricht. Bei der statischen Untersuchungsmethode nach Humphrey befinden sich die Reize an einer festgelegten Position in dem zu untersuchenden Gesichtsfeld und die Lichtintensität der Reize wird durch einen Computeralgorithmus variiert. Die Testperson drückt einen Knopf, wenn der Stimulus wahrgenommen wird, während die entsprechende Leuchtdichte notiert wird. Im Dossier wurden die Ergebnisse gemittelt über beide Augen für die Lichtempfindlichkeitsgrenze in der Einheit Dezibel für den Bereich Makula und Fovea angegeben.</p> <p>MID Für die Beurteilung der Wirksamkeit der Behandlung wurde die Anzahl der Personen erfasst, die zu Jahr 1B/C eine Verbesserung um > 3 dB und eine Verschlechterung um > 5 dB im Schwellenwert der Lichtempfindlichkeitsgrenze im Bereich der Makula und Fovea, gemessen über die Humphrey-Perimetrie, erreicht hatten.</p> <p>Auswertung/Auswertungspopulation Die Auswertung beider Endpunkte erfolgte auf Basis der ITT-Population für den Datenschnitt 16.07.2015.</p> <p>Erhebungszeitpunkte Zu Baseline, Monate 1B/C, 3B/C und 6B/C sowie Jahr 1B/C.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung kann nicht nachvollzogen werden. Unklar ist, warum sich konkrete Angaben zur Durchführung erst vereinzelt in Studienbericht und Herstellermodul M4 finden. Da es sich um eine unverblindete Studie handelt und gemäß Dossier auch keine verblindete Auswertung vorgesehen war, sind Angaben zur konkreten Testsituation notwendig. Dies trifft insbesondere auf multizentrische Studien sowie auf Studien mit Teilnahme von Kindern zu, da eine erfolgreiche Durchführung des Tests sowohl von der Motivation als auch vom Verstehen der Testsituation und der Handlungsanweisung abhängen kann. Angaben, inwieweit das Studienpersonal in der Durchführung des Tests geschult war, konnten ebenfalls nicht identifiziert werden. Unklar bleibt auch, ob eine Reihenfolge bei der Erhebung der primären, sekundären und explorativen Endpunkte vorgesehen war und auch ob mehrere Wiederholungen des Tests, ähnlich wie zu Baseline, auch zu anderen Visiten möglich war.</p> <p>Validität und Patientenrelevanz Bei beiden Perimetrie-Arten handelt es sich weit verbreitete Methoden zur Messung des Gesichtsfeldes, die der pU mit Beispielen unterlegt. Allerdings werden keine Validierungsstudien zu psychometrischen Eigenschaften bei einer Netzhautdystrophie für beide Arten der Perimetrie vorgelegt. In einer orientierenden Recherche konnte eine Publikation von Patel et al. (2011) [18] identifiziert werden, die sich mit der Reliabilität, Wiederholbarkeit und Durchführbarkeit bei Kindern beschäftigt, da der Erfolg auch von der Mitarbeit des Kindes abhängig sei. Der Anteil an 5- bis 8-Jährigen in der Studie war unter den Studienteilnehmenden (n = 154) mit 60,4 % am höchsten. Alle Teilnehmenden konnten die Untersuchungen</p>

Morbidität

abschließen. Bei beiden Testmethoden wurde bei den Teilnehmenden bei einer geringen Anzahl an Personen Fatigue festgestellt (Humphrey-Perimetrie: 5,2 %; Goldmann-Perimetrie: 8,4 %). Gemäß der Autoren konnten allerdings bei Kindern > 9 Jahre keine Symptome der Fatigue beobachtet werden. [18]

Insgesamt 24,2 % der Studienteilnehmenden empfanden den Perimetrie-Test nach Humphrey mittels einer 5-Punkte-Likert-Skala als am schwierigsten, während 63,3 % die Durchführung der Goldmann-Perimetrie als einfach empfanden. Wiederholbarkeit konnte für beide Perimetrie-Arten bei 27,9 % der 154 Personen untersucht und bei beiden Visiten (Baseline und 6 Monate später) nachgewiesen werden [18].

In der Studie wurden für die Humphrey-Perimetrie Reliabilitätskennzeichen mit definierten Grenzzahlen erhoben: Fixationsverluste (Patient/in gibt an, dass er/sie ein Stimulus im blinden Fleck sieht) ≥ 25 % oder falsch positive Ergebnisse ≥ 15 %. Beide Kriterien zusammen ergaben, dass 125 von 154 Messungen (81 %) als unreliabel klassifiziert werden können. Bei getrennter Betrachtung traf dies bei 42/154 (27 %) auf ein falsch-positives Ergebnis bzw. auf einen Fixationsverlust bei 117/154 (76 %) Untersuchungen zu [18]. Zusammenfassend zeigt die Studie von Patel et al. [18] eine gute Durchführbarkeit für beide Perimetrie-Arten bei Kindern. Die erhobenen Reliabilitätskriterien bei der Perimetrie nach Humphrey zeigen Einschränkungen in der Zuverlässigkeit der Messungen bei Kindern an.

Angaben zur Änderungssensitivität oder MID konnten für die Perimetrie nach Goldmann nicht identifiziert werden. Zur MID für die Perimetrie nach Humphrey wurde eine Validierungsstudie vorgelegt.

MID für Perimetrie nach Humphrey

Die Auswahl der MID begründet der pU anhand einer Studie von Patino et. al. (2011) [19]. Innerhalb der Studie wurden von 3.175 Teilnehmenden Gesichtsfeldmessungen mittels Humphrey-Perimetrie im Abstand von 4 Jahren durchgeführt. Ziel war es Änderungen in den Fragebögen zur Lebensqualität (NEI VFQ-25, SF-12) anhand von Änderungen im Gesichtsfeld (Humphrey Perimetrie) und Sehschärfe zu untersuchen. Die MID wurde ankerbasiert bestimmt.

Der NEI VFQ-25 ist ein Fragebogen zur konkreten Erfassung der alltagsrelevanten Einbußen des Sehvermögens für die Betroffenen, während es sich beim SF-12 um einen generischen Fragebogen handelt. Die Rekrutierung der Studienpopulation erfolgte bevölkerungsbezogen (Zensus). Die Studie schloss ausschließlich lateinamerikanische Personen im Alter von 40 Jahren oder älter ein. [19]

Für die Gesichtsfeldmessung wurden die Ergebnisse im Vergleich zu einer Referenzpopulation als mittlere Abweichung in Dezibel berichtet. Die für die Gesichtsfeldmessung verwendete intervallskalierte Skala wurde in eine ordinalskalierte Skala mit 2dB-Abständen umgewandelt. Als Begründung wurde die fehlende Interpretierbarkeit der Stärke einer Änderung im Gesichtsfeld auf Populationsebene angegeben. Zwar wurden sowohl intervall- als auch ordinalskalierte Ergebnisse berechnet, in der Studie werden nur die ordinalskalierten Daten angezeigt. [19]

Für den SF-12 und der Messung des Gesichtsfeldes konnte keine Assoziation festgestellt werden. Eine Verringerung im Gesichtsfeld von 5 dB bzw. eine Verbesserung um 3 dB war assoziiert mit einer Änderung im Score des NEI VFQ-25, wobei die MID für den NEI VFQ-25 abhängig war von der Schwere des Gesichtsfeldverlustes bzw. der Einschränkungen der Sehkraft zu Baseline.

Da das Ziel der Studie die Ermittlung einer MID für einen Fragebogen zur Bestimmung der Lebensqualität war (wie der NEI- VFQ-25), ist es unklar, inwieweit eine umgekehrte Bestimmung der MID durch den Fragebogen für das Gesichtsfeld geeignet ist. Anders als in Studie 301 wurde bei der Auswertung in der Validierungsstudie bezug auf eine Referenzpopulation genommen. Hinzu kommen Unterschiede in der Studienpopulation im Hinblick auf Alter (Studie 301: 4–44 Jahre; Patino et. al.: > 40 Jahre) und Grunderkrankung (Studie 301: Netzhautdystrophie; Patino et. al.: diabetische Retinopathie (90,4 %), Katarakt (16,2 %), AMD (34,3 %), Glaukom (41,0 %)) [19].

Morbidität	
	<p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p> <p>Die Messung des Gesichtsfeldes wird als patientenrelevant angesehen, allerdings ist aufgrund der künstlichen Messsituation beider Tests die Patientenrelevanz unklar.</p> <p>Zusammenfassung Die Endpunkte werden in der Nutzenbewertung nicht berücksichtigt. In der Gesamtschau sind die Informationen zur Operationalisierung nicht ausreichend und die Durchführung/Standardisierung des Tests ist nicht nachvollziehbar. In Anbetracht des offenen Studiendesigns und der Teilnahme von Kindern ist eine standardisierte Durchführung essentiell, um valide und unverzerrte Messungen durchzuführen. Angaben zu Reliabilität und Änderungssensitivität konnten ebenfalls nicht identifiziert werden.</p>
Endpunkt	Sehschärfetest
Operationalisierung	<p>Für die Untersuchung der Sehschärfe wurden die Sehtafel der Early Treatment of Diabetic Retinopathy Study (ETDRS) und der HOTV-Test angewandt.</p> <p>Subjektive Tests konnten zu Baseline mehrmals am Tag bzw. auf mehrere Tage verteilt gemessen werden, bis akzeptable Reliabilitätswerte erreicht wurden. Bei Kindern sollte die Dauer der Testdurchführungen kürzer sein, dafür an mehreren Tagen stattfinden. Unklar bleibt, ob ein ähnliches Vorgehen zu den weiteren Studienvisiten vorgesehen war und welche Reliabilitätskriterien verwendet wurden.</p> <p>ETDRS-Sehtafel Die ETDRS-Sehtafel ist aus 14 Sehzeichenreihen mit jeweils 5 Buchstaben aufgebaut (insgesamt 70 Buchstaben), wobei die Größe der Buchstaben mit jeder Reihe abnimmt. Die Sehschärfe wurde für den Vergleich zwischen den Behandlungsgruppen in den Logarithmus des minimalen Auflösungswinkels (logMAR) umgerechnet. Ein abnehmender logMAR-Wert ist somit mit einer besseren Sehschärfe gleichzusetzen. Eine Verbesserung der Sehschärfe ist definiert als die Zunahme um ≥ 10 und ≥ 15 ETDRS-Buchstaben (Abnahme logMAR um $\geq 0,2$ und $\geq 0,3$).</p> <p>Der Test kann unter bestimmten Bedingungen an mehreren Tagen wiederholt werden, um konsistente Messungen zu Baseline zu erhalten. Im Studienbericht wird zudem angegeben, dass die best-korrigierte Sehkraft untersucht wurde.</p> <p>Zur Auswertung finden sich erst im Studienbericht Informationen. Die vom pU angegebene Auswertung, angepasst nach Holladay et al. (2004) [9], wurde ergänzt durch die Auswertung anhand Lange et al. (2009) [12]. Nach Studienbeginn wurden seitens der EMA und des DSMB Bedenken geäußert, dass eine Differenz von 1-Log-Einheiten der Wahrnehmung von Fingerzählen und Handbewegung zu einer Verzerrung in der Berechnung des Effektschätzers führen kann. Bei dem Vorgehen von Lange et al. wird der 1-Log-Einheiten-Schritt reduziert auf 0,3-Log-Einheiten [12]. Für „Off-Chart“-Analysen wird auf eine SOP verwiesen, die in den Studienunterlagen nicht identifiziert werden konnte.</p> <p>MID-Bestimmung Für die ETDRS-Sehtafel wurden post hoc im Dossier zu Jahr 1B/C die Anzahl an Personen mit einer Verbesserung von ≥ 10 bzw. ≥ 15 Buchstaben (Abnahme logMAR um $\geq 0,2$ und $\geq 0,3$ Einheiten) bzw. einer Verschlechterung um ≥ 10 und ≥ 15 ETDRS-Buchstaben in der Sehschärfe (Zunahme logMAR um $\geq 0,2$ und $\geq 0,3$ Einheiten) untersucht.</p>

Morbidität	
	<p>HOTV-Test Beim HOTV-Test werden die Buchstaben H, O, T, V verwendet, die um eine vertikale Achse zentriert sind und somit auch von Kindern leichter identifiziert werden können. Es handelt sich dabei um Optotypen (= Sehzeichen) gleicher Lesbarkeit mit 5 Optotypen in gleicher Reihe. Die Größe der Buchstaben nimmt mit jeder Reihe ab. Die Sehschärfe kann ebenfalls in logMAR-Werte umgewandelt werden. Angaben zur Operationalisierung konnten in den Studienunterlagen nicht identifiziert werden.</p> <p>Angaben zur Auswertung der Ergebnisse des HOTV-Tests als auch zur konkreten Durchführung des Tests in der Studie 301 konnten nicht identifiziert werden.</p> <p>Auswertung/Auswertepopulation Die Auswertung erfolgte auf Basis der ITT-Population für den Datenschnitt 16.07.2015.</p> <p>Erhebungszeitpunkte Baseline, Tage 30B/C, 90 B/C und 180B/C sowie Jahr 1B/C.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung des Endpunkts kann nicht nachvollzogen werden. So wird berichtet, dass die Teilnehmenden die Buchstaben auf der Sehtafel von einem bestimmten Abstand aus ablesen sollen. Konkrete Angaben dazu, wie groß der Abstand genau sein soll und bei welcher Beleuchtungsstärke der Test in beiden Behandlungsarmen durchgeführt werden sollte, fehlen. Auch fehlen Angaben dazu, ab welchem Alter eine der beiden Sehtafeln angewandt werden sollten. Insbesondere bei Kindern ist eine standardisierte Durchführung des Tests notwendig, da der Erfolg der Durchführung von der Motivation, vom Verstehen der Handlungsanweisung und der Untersuchungsbedingungen abhängig ist. Unklar ist ebenfalls, ob eine Reihenfolge bei der Erhebung der primären, sekundären und zusätzliche Endpunkte vorgesehen war.</p> <p>Validität <u>ETDRS-Sehtafel</u> Die Erhebung der Sehschärfe anhand der ETDRS-Sehtafel ist ein etabliertes Verfahren. Aus verschiedenen Untersuchungen liegen Angaben zur Reliabilität vor. Dabei zeigte sich bei Menschen mit verschiedenen Augenerkrankungen, dass bei Wiederholungsmessungen Unterschiede (Test-Retest-Reliabilität) im BCDVA-Score bis zu 5–10 Buchstaben auftreten können, die als Messungenauigkeiten bzw. Ausdruck von individuellen Tagesschwankungen der Sehschärfe interpretiert werden können und zufällige Fehler darstellen [1,2,22,28]. Dabei ist die Test-Retest-Reliabilität abhängig von der Sehschärfe [2]. Es wird davon ausgegangen, dass mit der ETDRS-Sehtafel Veränderungen ab 10 Buchstaben reliabel gemessen werden können [21]. Welcher Grad an Veränderung im Score als klinisch relevant angesehen werden kann, ist noch nicht abschließend geklärt. In klinischen Studien wird häufig eine Veränderung um 10–15 ETDRS-Buchstaben als Responsekriterium verwendet. [4,5].</p> <p><u>HOTV-Test</u> Es wurden vom pU keine Studien zur Validität und zur Herleitung der MID vorgelegt. In einer Querschnittsstudie zur Untersuchung der Vergleichbarkeit und Prüfbarkeit von Sehschärfetests wurden bei australischen Vorschulkindern im Alter von mindestens 24 Monaten (n = 1.746) der „Amblyopia Treatment Study (ATS) HOTV protocol“ verwendet. Bei Kindern im Alter ab 60 Monaten (n = 576) die auf logMAR basierende HOTV-Sehtafel mit der auf logMAR basierenden ETDRS-Sehtafel (CSV 1000) verglichen. Es wurde die Prüfbarkeit untersucht, definiert als die Fähigkeit den Sehschärfetest mit beiden Augen beim ersten Mal vollständig abzuschließen. Die Prüfbarkeit betrug mindestens 93 % beim ATS HOTV bei Kindern ab 42 Monaten (n = 352). Bei Kindern ab 60 Monaten war die Prüfbarkeit mit dem HOTV-Test im Vergleich zur ETDRS-Sehtafel höher (HOTV: 94 % vs. ETDRS: 59 %) [13]. In einem aktuellen Bericht der US Preventive Services Task Force (USPSTF) wird der HOTV-Test zum Screening von Kindern zwischen 5 Monaten und 6 Jahren zur Ermittlung der Sehschärfe empfohlen [7].</p>

Morbidität	
	<p>Der HOTV-Test erscheint als ein verbreitetes Instrument zur Messung der Sehschärfe bei Kindern, obwohl Informationen zu Validität (beispielsweise Konstruktvalidität), Änderungssensitivität und MID nicht identifiziert werden konnten.</p> <p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Die Sehschärfe wird als unmittelbar patientenrelevant angesehen.</p> <p>Zusammenfassung Der Einsatz der EDTRS-Sehtafel und eine MID von 10–15 Buchstaben wird als angemessen angesehen. Der Einsatz der HOTV-Sehtafel bei Kindern mit eingeschränkter Sehkraft scheint angemessen zu sein. Unklar bleibt, ab welchem Alter eine der beiden Sehtafeln eingesetzt wurde. Der Endpunkt wird in der Nutzenbewertung nicht berücksichtigt. In der Gesamtschau sind die Informationen zur Operationalisierung nicht ausreichend und die Durchführung/Standardisierung des Tests ist nicht nachvollziehbar. In Anbetracht des offenen Studiendesigns und der Teilnahme von Kindern ist eine standardisierte Durchführung essentiell, um valide und unverzerrte Messungen durchzuführen.</p>
Endpunkt	Kontrastempfindlichkeit
Operationalisierung	<p>Die Messung der Kontrastempfindlichkeit war a priori geplant, jedoch fehlen im Studienprotokoll/SAP jegliche Angaben zur Erhebung. Ein Handbuch, Manual bzw. eine SOP wurde ebenfalls nicht eingereicht. Konkrete Angaben zur Durchführung finden sich vereinzelt im Studienbericht. So wurden mindestens zwei Lampen (jede mit drei 60-Watt-Glühbirnen) eingesetzt, deren photopischer Bereich gemäß einer SOP verifiziert war. Zudem wurde eine Pelli-Robson-Kontrasttafel verwendet und eine Änderung um 0,3 log als klinisch relevant definiert. Im Studienbericht wurde darauf hingewiesen, dass dieser Test gemäß SOP nicht bei Kindern eingesetzt werden sollte, die keine Buchstaben erkennen können.</p> <p>Analysepopulation/Auswertung Die Auswertung erfolgte für den Datenschnitt 16.07.2015. Auswertestrategien konnten sowohl im SAP als auch im Studienprotokoll nicht identifiziert werden.</p> <p>Erhebungszeitpunkte Baseline, Tage 30B/C, 90B/C und 180B/C sowie Jahr 1B/C.</p>
Bewertung	<p>Operationalisierung Eine Beurteilung der Operationalisierung ist aufgrund der wenigen Angaben im Studienbericht nicht möglich. Es wird auf ein Handbuch bzw. Manual verwiesen, das in den eingereichten Unterlagen nicht identifiziert werden konnte.</p> <p>Validität Es konnten keine Angaben zu Validität und Auswertung identifiziert werden. Da auch konkrete Angaben zu Durchführung und Auswertung fehlen, kann eine Beurteilung der Validität nicht stattfinden.</p> <p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Die Messung der Kontrastempfindlichkeit wird als patientenrelevant angesehen.</p>

Morbidität	
	<p>Zusammenfassung Aufgrund der fehlenden Angaben zu Operationalisierung, Auswertestrategie und Validität wird der Endpunkt nicht in der Nutzenbewertung berücksichtigt.</p>
Endpunkt	Orientierungs- und Mobilitätsmessungen
Operationalisierung	<p>Es wurden, durch geschulte Personen in der Beurteilung von Orientierung und Mobilität, bei den Studienteilnehmenden im Wohnumfeld Orientierungs- und Mobilitätsmessungen durchgeführt, um eine Bewertung der Leistung in alltäglicher Umgebung vorzunehmen. Es war geplant deskriptiv Fallberichte zu erstellen. Weitere Angaben konnten nicht identifiziert werden.</p> <p>Analysepopulation/Auswertung Die Auswertung erfolgte für den Datenschnitt 16.07.2015.</p> <p>Erhebungszeitpunkte Baseline und Jahr 1B/C.</p>
Bewertung	<p>Operationalisierung Eine Beurteilung der Operationalisierung ist nicht möglich, da konkrete Angaben zu Testsituation, Standardisierung, Bewertung des Tests und der Schulung der Beurteiler nicht identifiziert werden konnten.</p> <p>Validität In den Studienunterlagen wird angegeben, dass eine zusammenfassende Darstellung nicht geplant war, sondern die Ergebnisse in Form von Fallberichten dargestellt werden sollten. Eine Beurteilung der Validität ist aufgrund unzureichender Informationen im Studienbericht zum Endpunkt nicht möglich.</p> <p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der <u>Verlängerung des Überlebens</u>, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Eine Erfassung der Orientierung und Mobilität in alltäglicher Umgebung des Patienten / der Patientin wird als patientenrelevant angesehen.</p> <p>Zusammenfassung Aufgrund fehlender Angaben zu Operationalisierung und Validität sowie der fehlenden zusammenfassenden Darstellung der Ergebnisse wird der Endpunkt nicht in der Nutzenbewertung herangezogen</p>

Abkürzungen: AMD: Altersbedingte Makuladegeneration; AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; BCDVA: Best Corrected Distance Visual Acuity; cd (s): Candela (-Sekunde); DSMB: Data and Safety Monitoring Board; EMA: European Medicines Agency; ETDRS: Early Treatment of Diabetic Retinopathy Study; FDA: U.S. Food and Drug Administration; FST: Vollfeld-Lichtempfindlichkeits-Schwellenwert-Test; ITT: Intention-to-Treat; LCA: Lebersche kongenitale Anomalie; logMAR: Logarithmus des minimalen Auflösungswinkels; MID: Minimal Important Difference; MLMT: Multi-Luminanz-Mobilitätstest; NEI VFQ-25: National Eye Institute Visual Functioning Questionnaire – 25; pU: pharmazeutischer Unternehmer; SD: Standardabweichung; SF-12: Short Form 12 (Fragebogen); SOP: Standard Operating Procedure; Tag/Monat/Jahr xB: Tag/Monat/Jahr x nach Behandlung des zweiten Auges in der Interventionsgruppe; Tag/Monat/Jahr xC: Tag/Monat/Jahr x nach Baseline in der Kontrollgruppe.

Tabelle 7: Beschreibung der Endpunkte der Kategorie Lebensqualität

Endpunktkategorie	
Endpunkt	PRO-Instrument zur Erfassung der visuellen Lebensqualität
Operationalisierung	<p>Die visuelle Lebensqualität wird mithilfe eines Fragebogens untersucht. Der pU gibt an, dass für die Studie 301 eine Neuentwicklung auf Basis des NEI-VFQ 25 durchgeführt wurde, um eine auch für Kinder altersgerechte Erfassung der Lebensqualität zu ermöglichen, da sich manche Fragen des NEI VFQ-25 an erwachsenen Patientinnen und Patienten orientieren (u. a. Fragen zu Autofahren und Arbeitsfähigkeit). Beim NEI VFQ-25 handelt es sich um ein etabliertes, reliables und validiertes Instrument zur Erfassung der visuellen Lebensqualität. Er besteht aus 25 Fragen sowie aus 11 optionalen Zusatzfragen. Die Antwortkategorien variieren zwischen 5–6 Antwortoptionen. Diese werden anschließend umgewandelt auf eine Skala von 0–100, wobei 100 keine Einschränkungen und ein normales Sehvermögen bedeutet.</p> <p>Der vom pU adaptierte Fragebogen besteht ebenfalls aus 25 Fragen, die von den Patienten/Patientinnen selbst oder einem Elternteil/Erziehungsberechtigten in einem Bezugszeitraum von einem Monat beantwortet werden sollen. Sofern notwendig, konnten die Fragen laut vorgelesen werden, wobei Vorgaben zur Vorleseart definiert wurden. Die Fragebögen wurden während der Studienvisite ausgefüllt. Es gab keine zeitliche Begrenzung zur Beantwortung der Fragebögen, wobei dies bis zum Ende der Studienvisite erfolgen sollte.</p> <p>Die Unterschiede zwischen den beiden Fragebögen gemäß Dossier Modul 4 sind zum einen, dass im vom pU entwickelten Fragebogen keine Unterteilung mehr zwischen den Abschnitten „General Health and Vision“, „Difficulties with Activities“ und „Responses to Vision Problems“ gibt. Bei der adaptierten Version wird gemäß pU der Fokus auf die Domäne „Difficulties with Activities“ gelegt. Zudem wird im vorherigen Fragebogen eine Skala zur Beantwortung der Fragen von 5 bzw. 6 Punkten herangezogen. Beim in der Studie 301 verwendeten Fragebogen wird für Items eine Skala von 0–10 eingesetzt und anschließend der Durchschnitt der Werte gebildet. Der Gesamtwert des Patienten / der Patientin bzw. des Elternteils/Erziehungsberichtigten werden nicht gemittelt sondern separat über die Zeit ermittelt.</p> <p>Analysepopulation/Auswertung Die Auswertung erfolgte anhand der ITT-Population für den Datenschnitt 16.07.2015.</p> <p>Erhebungszeitpunkte Baseline, Tage 30B/C, 90B/C und 180B/C sowie Jahr 1B/C.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nachvollziehbar.</p> <p>Validität Die Fragen der beiden Fragebögen unterscheiden sich erheblich in ihrer Formulierung. Es bleibt unklar, nach welchen Kriterien die Fragen im vom pU entwickelten Fragebogen ausgewählt bzw. entwickelt wurden. Zusätzlich wurden die Antwortoptionen und die Struktur des Fragebogens geändert. Vom pU angeführte Validierungsstudien beziehen sich auf den bereits etablierten NEI VFQ-25. Studien zum vom pU entwickelten Fragebogen konnten in den eingereichten Unterlagen nicht identifiziert werden. Stattdessen gibt der pU an, die Ergebnisse der Validierungsstudien des NEI VFQ-25 auf den neu entwickelten Fragebogen übertragen zu wollen. Für die vom pU angegebene MID von ≥ 4 wurden ebenfalls Studien zum ursprünglichen Fragebogen zitiert. Aufgrund der erheblichen Unterschiede scheint weder eine Übertragbarkeit der psychometrischen Eigenschaften noch der MID vom NEI VFQ-25 auf den vom pU entwickelten Fragebogen möglich.</p> <p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des</p>

Endpunktkategorie	
	<p>Überlebens, der Verringerung von Nebenwirkungen oder einer <u>Verbesserung der Lebensqualität</u>.“</p> <p>Die visuelle Lebensqualität aus Patientensicht wird als patientenrelevant angesehen.</p> <p>Zusammenfassung</p> <p>Im Dossier werden Validierungsstudien für den ursprünglichen Fragebogen vorgelegt. Aufgrund der oben beschriebenen Unsicherheiten, die sich insbesondere aus inhaltlichen Unterschieden in den Fragen, der geänderten Antwortoptionen und der Struktur ableiten, ist die Übertragbarkeit der Ergebnisse der zitierten Validierungsstudien des NEI VFQ-25 auf das neu entwickelte Instrument nicht möglich. Aus diesen Gründen wird der Endpunkt „Visuelle Lebensqualität“ nicht in der Nutzenbewertung berücksichtigt.</p>

Abkürzungen: AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; ITT: Intention-to-Treat; MID: Minimal Important Difference; NEI VFQ-25: National Eye Institute Visual Functioning Questionnaire – 25; PRO: Patient-Reported Outcome; pU: pharmazeutischer Unternehmer; Tag/Jahr xB: Tag/Jahr x nach Behandlung des zweiten Auges in der Interventionsgruppe; Tag/Jahr xC: Tag/Jahr x nach Baseline in der Kontrollgruppe.

Tabelle 8: Beschreibung der Endpunkte der Kategorie Sicherheit

Endpunktkategorie	
Endpunkt	Unerwünschte Ereignisse (UE)
Operationalisierung	<p>Unerwünschte Ereignisse wurden durch Befragung oder durch den Patienten / die Patientin selbst berichtet bzw. durch Beobachtung und/oder körperliche Untersuchung erfasst. Die UE wurden gemäß MedDRA (Version 14.0 oder höher) nach Systemorganklasse und Preferred Term kodiert. Für die Interventionsgruppe wurden UE ab der ersten Injektion und für die Kontrollgruppe ab dem ersten Tag der Baseline-Untersuchung erfasst.</p> <p>Als UE wurden alle unvorhergesehenen oder nachteiligen Ereignisse definiert, unabhängig ob sie in Zusammenhang mit der Prüfmedikation standen oder nicht. Auffällige Laborparameter wurden nur als UE berichtet, sofern sie Grad 2 oder höher eingestuft wurden oder falls diese nach erneutem Testen bestehen blieben. Sofern Leukozytose und Hyperglykämie während der Gabe von Glukokortikoiden (nach Absetzen der Glukokortikoide) auftraten, wurden diese nur als UE berichtet, sofern die UE als Grad 3 oder höher eingestuft wurden. Bereits vor Baseline vorhandene Beschwerden wurden nur als UE bei einer Verschlechterung und/oder bei häufigerem Auftreten berichtet.</p> <p>Als schwerwiegendes unerwünschtes Ereignis (SUE) waren UE definiert, die zum Tod, zu einem lebensbedrohlichen Ereignis, zu einem stationären oder einem verlängerten Krankenhausaufenthalt, zu einer langanhaltenden oder signifikanten Unfähigkeit oder erheblichen Beeinträchtigung der Ausführung von alltäglichen Aktivitäten oder zu einer angeborenen Anomalie bzw. einem Geburtsfehler führten.</p> <p>Eine Schweregradeinteilung wurde vom pU sofern möglich vorgenommen:</p> <ul style="list-style-type: none"> • Mild: Symptome sind für den Teilnehmenden kaum spürbar und haben keinen Einfluss auf seine/ihre Leistung oder Funktion. • Moderat: Symptome mit ausreichendem Schweregrad, sodass Teilnehmende Beschwerden haben bzw. in der Verrichtung von täglichen Aktivitäten eingeschränkt sind. Eine Behandlung der Symptome könnte notwendig sein. • Schwer: Symptome verursachen schwere Beschwerden. Eine Behandlung der Symptome und/oder Hospitalisierung kann erfolgen. • Lebensgefährlich: Das Auftreten eines Ereignisses kann zu unmittelbarer Todesgefahr führen. • Tod: Das Auftreten eines Ereignisses führt zum Tod.

Endpunktkategorie	
	<p>Gemäß Dossier Modul 4 werden UE als mild, moderat, schwer (Grad 3), lebensbedrohlich (Grad 4) oder mit Todesfolge beschrieben.</p> <p>Wenn möglich, sollte gemäß Studienprotokoll eine Schweregradeinteilung in Orientierung der WHO-Toxizitätsskala erfolgen. Vom pU wurden ergänzend Schweregrade für ophthalmologische UE in dieser Einteilung vorgenommen.</p> <p>Es konnten sowohl im Studienprotokoll als auch im SAP keine a priori definierten UE von besonderem Interesse identifiziert werden. Es wurden post hoc im Dossier Effektschätzer für die Sicherheitsendpunkte berechnet.</p> <p>Folgende Sicherheitsendpunkte wurden im Studienbericht berichtet:</p> <ul style="list-style-type: none"> • UE • Schwere UE • SUE • UE, die zum Therapieabbruch führten <p>Im Dossier Modul 4 wurden zusätzlich UE ab Grad 3 berichtet.</p> <p>Auswertung/Auswertungspopulation</p> <p>Die Auswertung erfolgte auf Basis der Sicherheitspopulation für den Datenschnitt 16.07.2015.</p> <p>Erhebungszeitpunkte</p> <p>Es wurden zu jeder Studienvisite UE bis Jahr 1B erfasst. Bei Studienabbruch wurden UE bis 30 Tage später erfasst.</p> <p>UE wurden in der Interventionsgruppe ab der ersten Injektion und in der Kontrollgruppe ab Baseline erfasst.</p>
Bewertung	<p>Operationalisierung</p> <p>Die Operationalisierung ist nur zum Teil nachvollziehbar. Kritisch anzumerken ist der kürzere Erfassungszeitraum der UE in der Interventionsgruppe, die nicht ab Baseline (wie in der Kontrollgruppe) sondern ab dem Tag der ersten Injektion gezählt wurden, wodurch UE in der Interventionsgruppe zwischen Screening und Injektion ins erste Auge in der Behandlungsgruppe (≤ 90 Tage zwischen Screening und Baseline und ≤ 90 Tage zwischen Baseline und erster Injektion) nicht erfasst werden konnten. Kritisch anzumerken ist ebenfalls, dass bei der Erfassung einzelner UE (Leukozytosen, Hyperglykämien) erst ab einem Schweregrad ≥ 3 während der Einnahme von Glukokortikoiden geplant war.</p> <p>Validität</p> <p>Das Auftreten von unerwünschten Ereignissen wird als patientenrelevant und die Erhebung als größtenteils valide angesehen.</p> <p>Eine Schweregradeinteilung wurde in Orientierung der WHO-Toxizitätsskala mit einer zusätzlichen Schweregradeinteilung von ophthalmologischen UE durchgeführt. Inwieweit diese Schweregradeinteilung valide ist, kann nicht abschließend untersucht werden, da Kriterien für die Erstellung nicht identifiziert werden konnten.</p> <p>Kritisch anzumerken ist der unterschiedliche Beginn der UE-Erfassung in beiden Behandlungsgruppen. In der Interventionsgruppe erfolgte die Erfassung anders als in der Kontrollgruppe (ab Baseline) erst ab der ersten Injektion ins Auge. Der Zeitraum zwischen Baseline und Injektion konnte bis zu 90 Tage dauern. Dies kann gegenüber des längeren Erfassungszeitraums der Kontrollgruppe in der Interventionsgruppe zu einer Verzerrung der Ergebnisse zugunsten der Interventionsgruppe führen.</p> <p>Zusätzlich kann es durch die Erfassung von zwei UE (Leukozytose und Hyperglykämie) ab einem bestimmten Schweregrad für den Zeitraum der Glukokortikoideinnahme in der Interventionsgruppe nur zu einer in geringem Maße unvollständigen Erhebung aller UE in der Interventionsgruppe führen.</p>

Endpunktkategorie	
	Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der <u>Verringerung von Nebenwirkungen</u> oder einer Verbesserung der Lebensqualität“

Abkürzungen: AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; MedDRA: Medical Dictionary for Regulatory Activities; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; (S)UE: (Schwerwiegende/s) Unerwünschte/s Ereignis/se; WHO: World Health Organization.

Nach Betrachtung der Patientenrelevanz, Operationalisierung und Validität werden in dieser Nutzenbewertung die Ergebnisse der folgenden Endpunkte berücksichtigt:

- Multi-Luminanz-Mobilitätstest (MLMT), primärer Endpunkt
- Unerwünschte Ereignisse (UE)

Tabelle 9: Erhebungszeitpunkte der berücksichtigten Endpunkte in der Studie 301

Studienvisite Endpunkt	Baseline	Tag 0–8	Tag 14B ¹⁾	Tag 30B/C ¹⁾	Tag 90B/C ¹⁾	Tag 180B/C ¹⁾	Jahr 1B/C ¹⁾
MLMT	x	-	-	x	x	x	x ²⁾
Unerwünschte Ereignisse	x	x	x	x	x	x	x

¹⁾ Tag/Jahr xB: Tag/Jahr x nach Behandlung des zweiten Auges bzw. bei den Unerwünschten Ereignissen ab Behandlung des ersten Auges in der Interventionsgruppe; Tag/Jahr xC: Tag/Jahr x nach Baseline in der Kontrollgruppe.

²⁾ Primärer Endpunkt: Untersuchung der Änderung im MLMT-Score zu Jahr 1B/C im Vergleich zu Baseline bilateral gemessen zwischen den Behandlungsgruppen.

Abkürzungen: MLMT: Multi-Luminanz-Mobilitätstest.

2.5.3 Verzerrungspotential auf Studien und Endpunktebene

Tabelle 10: Verzerrungspotential der Studie 301

Studie	Adäquate Erzeugung der Randomisierungs- sequenz	Verdeckung der Gruppeneinteilung	Verblindung Patient/Patientin	Verblindung Studienpersonal	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
Studie 301	Ja ¹⁾²⁾	Ja ¹⁾	Nein ³⁾	Nein ³⁾⁴⁾	Nein	Nein ⁵⁾⁶⁾⁷⁾⁸⁾	Hoch

¹⁾ Die Randomisierung in die beiden Gruppen erfolgte zufällig durch ein unabhängiges Datenmanagementunternehmen nach der Screeningvisite nach Durchsicht der Screeningunterlagen. Die geplante Zuteilung wurden dem Studienpersonal erst bei den Untersuchungen zu Baseline bekannt gegeben. Es konnten keine detaillierten Informationen zum Randomisierungsprozess identifiziert werden. Aufgrund der zentralen Zuweisung durch ein externes Unternehmen und der Bekanntgabe nach der Baseline-Untersuchung, wird von einem angemessenen Verfahren zur verdeckten Zuteilung auf die beiden Behandlungsgruppen ausgegangen.

²⁾ Die Randomisierung erfolgte im Verhältnis 2:1 in Form einer Block-Randomisierung (Blockgröße von 3), stratifiziert nach Alter der Patienten (≥ 10 vs. < 10 Jahre) und Ergebnislevel des Mobilitätstests (bestehen bei ≥ 125 vs. < 125 Lux).

³⁾ Offenes Studiendesign.

⁴⁾ Die Ergebnisse des primären Endpunkts wurden verblindet ausgewertet.

⁵⁾ Gemäß Studienprotokoll war geplant weitere Personen bei vorzeitigem Ausscheiden von Testpersonen bzw. bei fehlender Applikation ins 2. Auge nachzurandomisieren. Gemäß Studienbericht sind zwei Personen vor der ersten Injektion ausgeschieden. Eine Aufnahme von zusätzlichen Personen erfolgte nicht.

- ⁶⁾ Es konnten erst im Studienbericht Angaben zur Auswahl des ersten Auges, das behandelt werden soll, identifiziert werden. So entschied das Prüfpersonal in Abhängigkeit der Sehschärfe und/oder Patientenpräferenz. Das erste Auge wies bei allen Teilnehmenden die schlechtere Sehschärfe auf; ausgenommen eine Person, bei der zu Baseline keine Unterschiede zwischen den Augen bestanden.
- ⁷⁾ Es haben sowohl in der Interventions- als auch in der Kontrollgruppe Testpersonen eine Begleitmedikation erhalten. Allerdings war der Anteil an Personen in der Interventionsgruppe, die Schmerzmittel (Anilide (Intervention vs. Kontrolle: 81,0 vs. 0,0 %), Propionsäure-Derivate (52,4 vs. 10,0 %)), Anticholinergika (47,6 vs. 0,0 %), Kortikosteroide + Antiinfektiva in Kombination (85,7 vs. 0,0 %) einnahmen, höher als in der Kontrollgruppe.
- ⁸⁾ Es finden sich Hinweise auf Unterschiede in der Studiendurchführung zwischen den Zentren, die in der Gesamtschau zu einer Verzerrung der Ergebnisse führen könnten: Die durchschnittliche Wartezeit bis zur ersten Injektion betrug im Studienzentrum CHOP 23,8 Tage (SD: 21,4), während diese im Studienzentrum IOWA länger war (44,7 Tage (SD: 30,7)). Zudem konnten mehr Abweichungen von der SOP/MOP im Studienzentrum IOWA im Vergleich zum Studienzentrum CHOP beobachtet werden. Darüber hinaus finden sich Hinweise auf eine ungleiche Verteilung der Kontrollpopulation auf die Studienzentren (CHOP: n = 8 (80,0 %) vs. IOWA: n = 2 (20,0 %)), während diese in der Interventionsgruppe ähnlich war (CHOP: n = 11 (52,4 %) vs. IOWA: n = 10 (47,6 %)).

Abkürzungen: CHOP: Children's Hospital of Philadelphia; IOWA: University of Iowa; MOP: Manual Operating Procedure; SD: Standardabweichung; SOP: Standard Operating Procedure.

Das Verzerrungspotential auf Studienebene wird als hoch eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 11 dargestellt.

Tabelle 11: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie 301

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
MLMT	Nein ¹⁾²⁾	Ja	Nein	Nein ⁵⁾	Hoch
Unerwünschte Ereignisse	Nein ¹⁾	Ja ³⁾	Nein ⁴⁾	Ja ⁶⁾⁷⁾⁸⁾	Hoch

¹⁾ Offenes Studiendesign.

²⁾ Beurteilung des Tests erfolgte verblindet.

³⁾ Sicherheitspopulation.

⁴⁾ Ergebnisse für UE \geq Grad 3 wurden nur zusammenfassend im Dossier berichtet. Unklar bleibt, inwieweit diese mit der Schweregradeinteilung im Studienbericht zusammenhängen, da sich die zusammenfassenden Ergebnisse unterschieden. Zudem konnten keine Angaben zu der zusätzlichen Schweregradeinteilung in Orientierung an der WHO-Toxizitätsskala identifiziert werden.

⁵⁾ Gemäß Dossier traten prozedurbedingte Abweichungen vom Verfahrenshandbuch beim MLMT am häufigsten in der Interventionsgruppe auf (Baseline vs. 1 Jahr: 3 vs. 2). In der Kontrollgruppe trat nur zu Baseline ein Ereignis auf. Angaben zur konkreten Anzahl an Personen konnte nicht identifiziert werden.

⁶⁾ Erfassung von bestimmten UE nur ab einem bestimmten Schweregrad in der Interventionsgruppe: Sofern Leukozytose und Hyperglykämie während der Gabe von Glukokortikoiden (18–30 Tage, je nach Zeitpunkt der zweiten Injektion) auftraten, wurden diese nur als UE berichtet, sofern sie als Grad 3 oder höher eingestuft wurden.

⁷⁾ Die UE-Erfassung in der Interventionsgruppe erfolgte erst ab der ersten Injektion, die durchschnittlich 34,3 Tage (SD: 27,9) nach Baseline stattfand. Dies führt zu einem kürzeren Erfassungszeitraums in der Behandlungsgruppe gegenüber der Kontrollgruppe, bei der die Erfassung bereits ab Baseline geplant war. Ähnliche Angaben zum Zeitraum Screening-Baseline für die Kontrollgruppe konnten nicht identifiziert werden.

⁸⁾ Angaben zur Studien- bzw. Beobachtungsdauer für beide Behandlungsgruppen konnten nicht identifiziert werden. Die Eignung bestimmter Auswerteverfahren der post hoc berechneten Effektschätzer (wie Relatives Risiko) bzw. inwieweit bei ungleichen Beobachtungszeiten geeigneter Verfahren wie Überlebenszeitanalysen angewandt werden können, kann nicht überprüft werden.

Abkürzungsverzeichnis: ITT: Intention-to-Treat; MLMT: Multi-Luminanz-Mobilitätstest; SD: Standardabweichung; UE: Unerwünschte/s Ereignis/se; WHO: World Health Organization.

2.5.4 Statistische Methoden

Relevante Änderungen in den Versionen des Statistischen Analyseplans

Es existieren 4 Versionen des SAP (Version 4: 19. August 2015; Version 3: 30. Juni 2015; Version 2: 10. März 2014; Version 1: 25. Januar 2013 bzw. gemäß Studienbericht der Studie LTFU/302 12. April 2013). Die Versionen wurden nach Einschluss der/des ersten Patienten/Patientin (15. November 2012) erstellt und stellen umfangreiche Änderungen des SAP dar, die nach Rücksprache mit der EMA (European Medicines Agency) bzw. der FDA (U.S. Food and Drug Administration) erfolgten. Version 4 wurde erst nach der Datenbanksperre vom 24. August 2015 erstellt, auf Grundlage eines Schreibens der FDA vom 21. Juli 2015.

Relevante Änderungen für die in der Nutzenbewertung dargestellten Endpunkte in Version 4 (19. August 2015): Ab Version 4 sollten die statistischen Analysen der primären Population auf Grundlage aller randomisierten Personen erfolgen. In Version 3 war die ITT-Population definiert als alle randomisierten Personen, die nicht ausgeschieden sind vor Bekanntgabe der Behandlungszuteilung an die Teilnehmenden, das Studienpersonal oder an den medizinischen Monitor, und beinhaltete alle Personen, die entweder eine bzw. beide Injektionen erhalten haben.

Relevante Analysepopulationen

Intention-to-Treat (ITT-) Population (Intervention: N = 21; Kontrolle: N = 10):

Alle Personen, die randomisiert und in die Studie aufgenommen wurden.

Modified ITT (mITT-) Population (Intervention: N = 20; Kontrolle: N = 9):

Alle randomisierten Personen, die nicht ausgeschieden sind bevor die folgenden Personen die Behandlungszuteilung kannten: Patient/in, Elternteil, primäre/r Prüfärztin/-arzt, medizinischer Monitor.

Sicherheitspopulation (Intervention: N = 20; Kontrolle: N = 9):

Alle Personen der Interventionsgruppe, die die Prüfmedikation in beide Augen erhalten haben.

Für die Kontrollgruppe wird ein ähnlicher Passus beschrieben wie für die mITT-Population: alle Personen der Kontrollgruppe, die nicht ausgeschieden sind bevor die folgenden Personen die Behandlungszuteilung kannten: Patient/in, Elternteil, primäre/r Prüfärztin/-arzt, medizinischer Monitor.

Gemäß Dossier Modul 4 waren mITT- und Sicherheitspopulation identisch.

Per Protocol (PP-) Population (Intervention: N = 19; Kontrolle: N = 9):

Alle Personen der ITT-Population, die 1. alle Ein-/Ausschlusskriterien erfüllen und 2. nicht ausgeschieden sind bevor der/die Patient/in, Elternteil, primäre/r Prüfärztin/-arzt, medizinischer Monitor die Zuteilung der randomisierten Behandlungszuteilung kannten. Patienten / Patientinnen, die die Prüfmedikation in der Behandlungsgruppe nicht in beide Augen erhalten haben, wurden ausgeschlossen.

Fallzahlplanung

Für die Fallzahlplanung führt der pU die Studie 102 auf, bei der in ein Auge die Prüfmedikation appliziert wurde. In der Studie wurde eine ähnliche Bewertung wie beim MLMT der Studie 301 herangezogen. Bei 80 % der Teilnehmenden trat eine Verbesserung um mindestens einen Punkt 1 Jahr nach Baseline nach Erhalt der Prüfmedikation auf. Diese Änderungen wurde vom pU als

klinisch bedeutsam definiert. Gemäß pU wurde für die Kontrollgruppe aufgrund der langsamen Progression der Erkrankung eine Änderung von 0 geschätzt.

Bei 16 Personen in der Interventionsgruppe und 8 Personen in der Kontrollgruppe (alle Werte bei 0) liegt die geschätzte Power, um eine wie vom pU definierte klinisch bedeutsame Änderung mittels Wilcoxon-Rangsummen-Tests mit exaktem p-Wert zu detektieren, bei 100 %. Für die Kontrollgruppe wurde bei Simulationsstudien eine Power von 94 % bei einer Wahrscheinlichkeit von 20 % für eine Verbesserung im Score um einen Punkt geschätzt.

Somit war eine Fallzahl von 24 Personen definiert. Zusätzlich wurde beschrieben, dass mindestens 10 % oder 3 Personen (2 in die Interventions-, 1 in die Kontrollgruppe) mehr randomisiert werden konnten, sofern Daten für die Bewertung von einem/einer in die Studie randomisierten Patienten/Patientin nicht zur Verfügung standen.

Es war geplant die Rekrutierung zu stoppen, sobald 18 Personen die Injektion ins zweite Auge erhalten haben und 9 Personen bereits in die Kontrollgruppe aufgenommen wurden. Aufgrund der Stratifizierungsfaktoren war es zudem erforderlich insgesamt 4 Personen zusätzlich zu rekrutieren, wodurch sich die Gesamtzahl auf 31 Personen erhöht.

Statistische Methoden

Für die statistische Analyse wurde eine externe Firma beauftragt. Neben dem SAP der Studie 301 und dem Dossier Modul 4 finden sich zusätzliche Informationen zu Analyseverfahren in einem SAP der vom pU beauftragten externen Firma [11].

Alle Daten wurden deskriptiv für jede Behandlungsgruppe dargestellt. Kategoriale Variablen wurden mittels Häufigkeiten, Prozent- und kontinuierliche Angaben wurden mittels Mittelwert (MW), Median, Standardabweichung (SD), relevante Perzentile, Minimal- und Maximalwerte zusammengefasst.

Für die Wirksamkeitsanalysen in der Behandlungsgruppe wurde ab dem Tag der zweiten Injektion (Tag 0B) gerechnet. Für die Sicherheitsanalysen der Interventionsgruppe wurde der Tag der ersten Injektion (Tag 0A) herangezogen. Für Analysen der Kontrollgruppe wurde stets der erste Tag der Baseline-Untersuchungen verwendet.

Die Randomisierung erfolgte stratifiziert nach Alter (≥ 10 vs. < 10 Jahre) und Ergebnislevel des MLMT (bestehen bei ≥ 125 vs. < 125 Lux), während die statistischen Analysen gemäß pU aufgrund der geringen Fallzahl ($n = 31$) und damit verbundener möglicher Konvergenzproblemen nicht stratifiziert durchgeführt wurden.

Analyse primärer Endpunkt

Die primäre Hypothese untersuchte gemäß SAP, inwiefern das bilateral applizierte Prüfpräparat bei Erwachsenen und Kindern ab 3 Jahren die Fähigkeit zu Navigieren (gemessen anhand des MLMT) verbesserte. Hierfür wurde für die Analyse des primären Endpunkts die Änderung im MLMT-Score bilateral nach einem Jahr (Visite Jahr 1B/C) im Vergleich zu Baseline zwischen den beiden Behandlungsgruppen untersucht.

Die Analyse beruht auf dem Wilcoxon-Rangsummen-Test. Die Berechnung der Teststatistik des Wilcoxon-Rangsummen-Tests basiert auf Rängen, die den Messwerten zugeordnet werden. Hatten Beobachtungen den gleichen Wert (ties/gebunden), so wurde der durchschnittliche Rang

verwendet. Da man viele gebundene Werte vermutet (wie beispielsweise eine hohe Anzahl des Wertes 0), wurde die exakte Methode zur Berechnung des p-Wertes als angemessen angesehen. Neben der statistischen Analyse mit dem Wilcoxon-Rangsummen-Test wurde gemäß Dossier zusätzlich die Differenz der mittleren Veränderungen berechnet. Das zugehörige Konfidenzintervall wurde mittels eines gemischten Modells berechnet, mit Behandlung und Studienvisite als Terme. Es fehlen weitere Angaben zum Modell, wie beispielsweise Angaben zur Kovarianzstruktur. Angaben zu dem Modell konnten erst in einem externen Dokument und im Dossier, aber nicht im SAP der Studie 301 für den primären Endpunkt identifiziert werden.

Relevante Kriterien für den Umgang mit fehlenden Werten - primärer Endpunkt

Bei Personen, die aus der Studie ausschieden, wurde angenommen, dass sie sich nicht verbesserten. Diesen Personen wurde in der Analyse eine Änderung von 0 zur Visite Jahr 1B/C zugewiesen.

Analysen zu Unerwünschten Ereignissen

Es wurden im Dossier post hoc Effektschätzer (Odds Ratio, Relatives Risiko und Risikodifferenz) sowie p-Werte mittels dem exakten Test nach Fischer berechnet. Das Konfidenzintervall wurde mit einer exakten Methode (Tail-Methode) berechnet durch Umkehren zweier separater einseitiger Tests, bei der die Größe des Tests höchstens $\alpha/2$ und das Konfidenzintervall mindestens $(1-\alpha)$ ist.

Gemäß SAP war eine deskriptive Analyse geplant.

2.6 Ergebnisse zum Zusatznutzen

2.6.1 Charakterisierung der Studienpopulation

Für die Studie 301 wurden 36 Personen gescreent und anschließend 31 Personen (ITT-Population – Interventionsgruppe N = 21, Kontrollgruppe N = 10) in die Studie aufgenommen (Tabelle 12). Alle Teilnehmenden der Kontrollgruppe, die die Studie zu Jahr 1 nach Baseline abschlossen (n = 9), erhielten anschließend die Prüfmedikation.

Gemäß Dossier M4 lag die Studiendauer zwischen einem Jahr und einem Jahr und 4 Monaten in der Interventions- und der Kontrollgruppe. Konkrete Angaben zur Beobachtungsdauer von Baseline bis zur Visite Jahr 1B/C bzw. zur gesamten Studiendauer sowie zum Zeitraum zwischen Screening und Baseline für beide Behandlungsgruppen konnten nicht identifiziert werden.

Angaben zu Kriterien der Auswahl des zuerst behandelten Auges finden sich erst im Studienbericht. Das erste Auge, das behandelt werden sollte, wurde in Abhängigkeit der Sehschärfe und/oder nach Patientenpräferenz ausgewählt. Bis auf eine Person, bei der die Sehschärfe zwischen den Augen ähnlich war, wurde das Auge als erstes operiert, das die schlechtere Sehschärfe aufwies.

In der Interventionsgruppe erhielt eine Person die Prüfmedikation nicht, da sie aufgrund einer Entscheidung des Prüfpersonals aus der Studie ausgeschlossen wurde. Bei dieser Person lag eine bestätigte schwere retinale Atrophie/Degeneration vor. Für die Interventionsgruppe waren somit für 20 (90 %) Personen Daten für die Visite 1 Jahr nach Baseline vorhanden. In der Kontrollgruppe hat eine Person die Einverständniserklärung zurückgezogen, wodurch 9 von 10 Personen (90 %) die Studie zu Jahr 1 abgeschlossen haben. Beide Personen sind aus der Studie ausgeschieden, bevor sie die Zuteilung in eine der beiden Behandlungsgruppen erfahren haben (Tabelle 12).

Einen Hinweis auf schiefe Verteilung zeigt sich bei der Zeit zwischen Randomisierung bis zur ersten Injektion (MW (SD): 34,3 (27,9) Tage; Median (min; max): 22 (6; 78) Tage) und der Zeit zwischen erster und zweiter Injektion (MW (SD): 8,8 (2,6) Tage; Median (min; max): 7 (7; 14) Tage). Angaben zu Beobachtungsdauer, Studiendauer oder ähnliches konnte für beide Behandlungsgruppen nicht identifiziert werden.

Tabelle 12: Allgemeine Angaben zur Studie 301

Studie 301	Intervention: Voretigen Neparvovec n (%)	Kontrolle: Beobachtendes Abwarten n (%)
ITT-Population ¹⁾	21 (100)	10 (100)
mITT-Population ²⁾	20 (95,2)	9 (90,0)
Sicherheitspopulation ³⁾	20 (95,2)	9 (90,0)
PP-Population ⁴⁾⁵⁾	19 (90,5)	9 (90,0)
Studiendauer (Einschluss in die Studie / Screening bis Studienende)	k.A.	k.A.
Beobachtungsdauer (Baseline / Injektion 1./2. Auge bis prim. EP)	k.A.	k.A.
Zeit zwischen Screening- und Baseline-Visite	k.A.	k.A.
Anzahl Patienten zum Studienende (Jahr 1 nach Baseline)	20 (95,2)	9 (90,0)
Vorzeitiges Ausscheiden aus der Studie Vor Behandlung Nach Behandlung	1 (4,8) 0 (0,0)	1 (10,0) 0 (0,0)
Gründe für das vorzeitige Ausscheiden Entscheidung des Arztes Einverständniserklärung zurückgezogen	1 (4,8) 0 (0,0)	0 (0,0) 1 (10,0)
Applikation der Prüfmedikation in beide Augen	20 (95,2)	n.a.
Zeit von Randomisierung bis zur ersten Injektion in Tagen Mittelwert (SD) Median (min; max)	20 (95,2) 34,3 (27,9) 22 (6; 78)	n.a.
Zeit zwischen ersten und zweiten Injektion in Tagen Mittelwert (SD) Median (min; max)	20 (95,2) 8,8 (2,6) 7 (7; 14)	n.a.
Jegliche Protokollabweichungen/-verletzungen Abweichungen von der Vorgehensweise (SOP/MOP) Untersuchungen nicht im geforderten Zeitraum durchgeführt Abweichungen im Zusammenhang mit der Medikation ⁶⁾ Eine Untersuchung wurde nicht durchgeführt Einschlusskriterien nicht erfüllt	17 (81,0) 14 (66,7) 5 (23,8) 5 (23,8) 4 (19,0) 1 (4,8)	5 (50,0) 2 (20,0) 2 (20,0) 1 (10,0) 0 (0,0) 0 (0,0)
<i>Studienzentren</i> Children's Hospital of Philadelphia (CHOP), USA University of Iowa (IOWA), USA	11 (52,4) 10 (47,6)	8 (80,0) 2 (20,0)
<i>Alter</i> Personen ≥ 10 Jahre Personen < 10 Jahre	12 (57,1) 9 (42,9)	6 (60,0) 4 (40,0)
<i>Screeningergebnis des MLMT⁷⁾</i> Bestehen bei < 125 Lux Bestehen bei ≥ 125 Lux	12 (57,1) 9 (42,9)	4 (40,0) 6 (60,0)

- 1) Alle Personen, die in eine der beiden Behandlungsgruppen randomisiert wurden.
- 2) Alle randomisierten Personen, die nicht ausgeschieden sind bevor die folgenden Personen die Behandlungszuteilung kannten: Patient/in, Elternteil, primäre/r Prüfärzt/in, medizinischer Monitor; mITT- und Sicherheitspopulation waren identisch.
- 3) In der Sicherheitspopulation sind alle Personen der Interventionsgruppe, die die Prüfmedikation in beide Augen erhalten haben. Für die Kontrollgruppe wird ein ähnlicher Passus beschrieben wie für die mITT-Population: alle Personen der Kontrollgruppe, die nicht ausgeschieden sind bevor die folgenden Personen die Behandlungszuteilung kannten: Patient/in, Elternteil, primäre/r Prüfärztin/-arzt, medizinischer Monitor.
- 4) In der PP-Population befinden sich alle Personen der ITT-Population, die 1. alle Ein-/Ausschlusskriterien erfüllen und 2. nicht ausgeschieden sind bevor Patient/in, Elternteil, primäre/r Prüfärztin/-arzt und der medizinische Monitor die Zuteilung der randomisierten Behandlungszuteilung kannten. Patienten/Patientinnen, die die Prüfmedikation in der Behandlungsgruppe nicht in beide Augen erhalten haben, wurden ausgeschlossen.
- 5) Eine Person hat zum Screening den MLMT bilateral bei 1 Lux bestanden, nicht aber während der beiden unilateralen Testsituationen. Dies wurde erst nach der Injektion des Prüfpräparats festgestellt. Die Person, obwohl sie nicht die Ein-/Ausschlusskriterien erfüllte, wurde weiter beobachtet und die Daten für die ITT- und mITT-Analyse, nicht aber für die PP-Analyse verwendet.
- 6) Abweichungen standen in Zusammenhang mit dem Prednisolon-/Prednison-Dosierungsregime.
- 7) Es wurde das Auge mit dem schlechtesten Testergebnis verwendet.

Abkürzungen: ITT: Intention-to-Treat; k.A.: keine Angaben; mITT: modified Intention-to-Treat; MLMT: Multi-Luminanz-Mobilitätstest; MOP: Manual Operating Procedure; n.a.: nicht anwendbar; PP: Per Protocol; prim. EP: primärer Endpunkt; SD: Standardabweichung; SOP: Standard Operating Procedure.

Es traten während der Studie prozedurbedingte Protokollabweichungen auf, die in der Regel auf Abweichungen vom Verfahrenshandbuch und/oder der Prozedur bei der Durchführung der Seh- und Netzhautfunktionstests und nicht auf Abweichungen vom klinischen Protokoll zurückzuführen waren. Gemäß Dossier traten prozedurbedingte Abweichungen vom Verfahrenshandbuch beim MLMT am häufigsten in der Interventionsgruppe auf (Baseline vs. 1 Jahr: 3 vs. 2 Abweichungen). In der Kontrollgruppe trat nur zu Baseline eine prozedurbedingte Abweichung auf. Gründe für die höhere Anzahl an Abweichungen beim MLMT waren, dass der Test für bestimmte Beleuchtungsstärken nur mit dem linken und dem rechten Auge, aber nicht mit beiden Augen durchgeführt wurde. Weitere Gründe waren die Doppelung eines Tests sowie die mangelnde Verwertbarkeit einer Videoaufzeichnung. In diesen Fällen wurde der Test wiederholt.

Das Gesamtvolumen, das ins erste Auge injiziert wurde, betrug durchschnittlich 310,0 µl (SD: 44,7) und im Median 300 µl (min; max: 300; 500); beim zweiten Auge waren es durchschnittlich 300,0 µl (SD: 0,0) und im Median 300 µl (min; max: 300; 300). Die Anzahl an Injektionen betrug ins erste Auge durchschnittlich 1,4 (SD: 0,5) und im Median 1 (min; max: 1; 2), ins zweite Auge 1,2 (SD: 0,5) und im Median 1 (min; max: 1; 3).

Insgesamt haben 90 % der Personen im Behandlungsarm die Prüfmedikation mit dem zulassungsrelevanten Volumen von 300 µl erhalten. Gemäß Dossier wurde bei zwei Personen eine Reduktion des Volumens der ersten Injektion (rechtes Auge) auf $1,0 \times 10^{11}$ Vg (200 µl) bzw. $1,25 \times 10^{11}$ Vg (250 µl) vorgenommen. Der Grund war, dass bei der Injektion der ersten 100 bzw. 50 µl keine subretinale Blase beobachtet werden konnte, sodass eine zweite Injektionsstelle für die verbleibenden 200 bzw. 250 µl erforderlich war.

Bei einer Person wurden insgesamt 500 µl injiziert, davon 300 µl in den subretinalen Raum des ersten Auges. Ein Reflux von ca. 200 µl wurde in dem Bereich des Glaskörpers beobachtet, der vor der erneuten Injektion entfernt wurde.

Unterschiede zwischen den Studienzentren

In die Interventionsgruppe wurden gleichmäßig Patientinnen und Patienten zwischen den beiden Behandlungszentren eingeschlossen (CHOP vs. IOWA: 52 vs. 48 %), während im Studienzentrum CHOP mehr Personen in die Kontrollgruppe eingeschlossen wurden (80 vs. 20 %).

Unterschiede gab es in der Dauer zwischen dem Zeitpunkt der Randomisierung und der ersten Injektion zwischen den Studienzentren. Beim Studienzentrum CHOP betrug die Wartezeit durchschnittlich 23,8 Tage (SD: 21,4), im Studienzentrum IOWA betrug die Wartezeit 44,7 Tage (SD: 30,7). Unterschiede in den Randomisierungsstrata innerhalb der Studienzentren wurden nicht beobachtet. Die Zeit zwischen erster und zweiter Injektion war vergleichbar (CHOP vs. IOWA: 9,8 Tage (SD: 2,6) vs. 7,7 Tage (SD: 2,2)).

Im Studienbericht wurde berichtet, dass es vermehrt zu Abweichungen vom SOP/MOP im Studienzentrum IOWA im Vergleich zum Studienzentrum CHOP kam. Ein intensives Training/Management durch den Sponsor wurde durchgeführt.

Charakterisierung der Studienpopulation

In Tabelle 13 sind die Charakteristika der Studienpopulation dargestellt.

Eine vorhergehende Medikation haben 57 % der Personen in der Interventions- und 80 % der Personen in der Kontrollgruppe erhalten. Am häufigsten in der Interventionsgruppe wurden Vitaminkombinationspräparate (24 %) eingesetzt (Kontrollgruppe: 10,0%), während in der Kontrollgruppe andere Mittel, die den Lipidstoffwechsel beeinflussen (40,0 %) verabreicht wurden (Interventionsgruppe: 4,8 %).

In der medizinischen Vorgeschichte traten in beiden Behandlungsgruppen keine Hornhauttrübung, Glaukoma, Keratokonus oder okuläre Anomalien auf.

Es konnten Imbalancen in Bezug auf die Begleitmedikation beobachtet werden. Der Anteil an Personen in der Interventionsgruppe, die Schmerzmittel (Anilide (Intervention vs. Kontrolle: 81,0 vs. 0,0 %), Propionsäure-Derivate (52,4 vs. 10,0 %)), Anticholinergika (47,6 vs. 0,0 %), Kortikosteroide + Antiinfektiva in Kombination (85,7 vs. 0,0 %) war höher als in der Kontrollgruppe.

Im Studienzentrum IOWA wurden in die Kontrollgruppe Personen aufgenommen, die durchschnittlich 29,5 Jahre (SD: 2,1) alt waren (Median 30 (min; max: 28; 31), während in der Interventionsgruppe jüngere Personen teilnahmen (MW (SD): 17,0 (14,7); Median (min; max): 11 (4; 44)). Weitere Imbalancen konnten sowohl für das Studienzentrum IOWA als auch das Studienzentrum CHOP nicht identifiziert werden.

Tabelle 13: Charakterisierung der Studienpopulation der Studie 301 (ITT-Population)

Studie	Intervention: Voretigen Neparvovec (N = 21)	Kontrolle: Beobachtendes Abwarten (N = 10)
<i>Alter zum Zeitpunkt der Randomisierung (Jahre)</i>		
Mittelwert (SD)	14,7 (11,8)	15,9 (9,5)
Median (min; max)	11 (4; 44)	14 (4; 31)
<i>Alter zum Zeitpunkt der Diagnosestellung</i>	k.A.	k.A.
<i>Geschlecht, n (%)</i>		
männlich	9 (42,9)	4 (40,0)
weiblich	12 (57,1)	6 (60,0)

Studie	Intervention: Voretigen Neparvovec (N = 21)	Kontrolle: Beobachtendes Abwarten (N = 10)
<i>Ethnie (genetisch), n (%)</i>		
europäisch	14 (66,7)	7 (70,0)
asiatisch	3 (14,3)	2 (20,0)
indigen	2 (9,5)	1 (10,0)
afroamerikanisch	2 (9,5)	0 (0,0)
<i>Region, n (%)</i>		
Europa	1 (4,8)	4 (40,0)
Nordamerika	18 (86)	6 (60,0)
Lateinamerika	1 (4,8)	0 (0,0)
Asien/Pazifik	1 (4,8)	0 (0,0)
<i>Ethnie, n (%)</i>		
nicht spanisch oder lateinamerikanisch	16 (76,0)	9 (90,0)
spanisch oder lateinamerikanisch	5 (24,0)	1 (10,0)
<i>Verteilung der Personen, bei denen der MLMT bilateral bei der niedrigsten Beleuchtungsstärke noch bestanden wurde zu Baseline, n (%)</i>		
1	0 (0,0)	0 (0,0)
4	4 (19,0)	1 (10,0)
10	5 (23,8)	2 (20,0)
50	7 (33,3)	5 (50,0)
125	3 (14,3)	1 (10,0)
250	0 (0,0)	0 (0,0)
400	0 (0,0)	0 (0,0)
> 400 ¹⁾	2 (9,5)	1 (10,0)
<i>Mindestens eine Begleitmedikation²⁾, n (%)</i>	20 (95,2)	9 (90,0)
Anilide	17 (81,0)	3 (30,0)
Paracetamol	15 (71,4)	1 (10,0)
Anticholinergika	10 (47,6)	0 (0,0)
Cyclopentolat Hydrochlorid	10 (47,6)	0 (0,0)
Kortikosteroide + Antiinfektiva in Kombination	18 (85,7)	0 (0,0)
Dexamethason + Tobramycin ³⁾	17 (81,0)	0 (0,0)
Corticosteroids, plain	6 (28,6)	0 (0,0)
Prednisolone acetate	6 (28,6)	0 (0,0)
Andere Antiinfektiva	9 (42,9)	5 (50,0)
Moxifloxacin Hydrochlorid	9 (42,9)	5 (50,0)
Propionsäure-Derivate	11 (52,4)	1 (10,0)
Ibuprofen	9 (42,9)	1 (10,0)
Andere Mittel, die den Lipidstoffwechsel beeinflussen	0 (0,0)	3 (30,0)
Omega-3-Fettsäuren inkl. andere Ester und Säuren	0 (0,0)	3 (30,0)
<i>Medizinische Vorgeschichte: jeglicher Befund⁴⁾, n (%)</i>	19 (90,5)	8 (80,0)
Allergie	9 (42,9)	5 (50,0)
Hals-Nasen-Ohren	10 (47,6)	4 (40,0)
Gastrointestinaltrakt	3 (14,3)	2 (20,0)
Urogenitales System	4 (19,0)	3 (30,0)
Integumentsystem	7 (33,3)	3 (30,0)
Muskuloskeletales System	5 (23,8)	3 (30,0)
Neurologisches System	5 (23,8)	4 (40,0)
Andere	7 (33,3)	2 (10,0)

Studie	Intervention: Voretigen Neparvovec (N = 21)		Kontrolle: Beobachtendes Abwarten (N = 10)	
	Erstes Auge	Zweites Auge	Erstes Auge	Zweites Auge
Ophthalmologische Vorgeschichte				
<i>Jeglicher Befund</i>	21 (100)	21 (100)	10 (100)	10 (100)
Katarakt	1 (4,8)	1 (4,8)	1 (10,0)	1 (10,0)
Linsentrübung	1 (4,8)	1 (4,8)	2 (20,0)	1 (10,0)
Nystagmus	21 (100)	21 (100)	10 (100)	10 (100)
Retinale Anomalien	21 (100)	21 (100)	10 (100)	10 (100)
Strabismus	8 (25,8)	7 (33,0)	5 (50,0)	4 (20,0)
Andere	3 (14,3)	1 (4,8)	2 (20,0)	2 (20,0)

¹⁾ Personen, die den Test bei 400 Lux nicht bestanden haben.

²⁾ Es wurde die Begleitmedikation ab einer Häufigkeit von $\geq 30\%$ in einer der beiden Behandlungsgruppen dargestellt.

³⁾ Im Studienbericht: Tobradex.

⁴⁾ Es wurde die medizinische Vorgeschichte ab einer Häufigkeit von $\geq 20\%$ in einer der beiden Behandlungsgruppen dargestellt.

Abkürzungen: ITT: Intention-to-Treat; k.A.: keine Angaben; MLMT: Multi-Luminanz-Mobilitätstest; SD: Standardabweichung.

2.6.2 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserhebungen erfasst. Es traten in der Studie keine Todesfälle auf.

2.6.3 Morbidität

Multi-Luminanz-Mobilitätstest (MLMT)

Die Ergebnisse für den primären Endpunkt sind in Tabelle 14 dargestellt. Für zwei Personen (je eine Person in der Interventions- bzw. Kontrollgruppe) waren nur Baseline-Werte vorhanden, da diese Personen vorzeitig vor Administration der Prüfmedikation aus der Studie ausschieden. Für beide Personen wurde ein Änderungswert von 0 zu Jahr 1 imputiert. Die Ergebnisse zu den Visiten zu den Tagen 30B/C, 90B/C und 180B/C sind konsistent mit den Ergebnissen zu Jahr 1B/C.

Tabelle 14: Änderung im MLMT bilateral gemessen, Jahr 1B/C im Vergleich zu Baseline (ITT-Population)

MLMT ¹⁾	Intervention: Voretigen Neparvovec (N = 21)	Kontrolle: Beobachtendes Abwarten (N = 10)
<i>Baseline</i>		
n	21	10
Score: Mittelwert (SD)	3,1 (1,7)	2,9 (1,6)
Score: Median (min; max)	k.A.	k.A.
<i>Jahr 1²⁾</i>		
n ³⁾	21	10
Score: Mittelwert (SD)	4,9 (2,0)	3,2 (1,8)
Änderung zu Baseline: Mittelwert (SD)	1,8 (1,1)	0,2 (1,0)
Median (min; max)	2 (0; 4)	0 (-1; 2)
Exakter p-Wert ⁴⁾		0,001 ⁵⁾
Änderung Intervention - Kontrolle [95%-KI] ⁶⁾		1,6 [0,7; 2,4]

- 1) Ein höherer Wert im Mobilitätsscore bedeutet eine Verbesserung im MLMT.
- 2) Primärer Endpunkt.
- 3) Fehlende Werte zu Jahr 1 von Personen, die vor Applikation der Prüfmedikation ausgeschieden sind, wurden mit 0 imputiert.
- 4) Zur Untersuchung der Änderung zu Jahr 1 im Vergleich zu Baseline zwischen den Behandlungsgruppen wurde ein Wilcoxon-Rangsummen-Test mit exaktem zweiseitigem p-Wert verwendet.
- 5) Im Dossier wird ein p-Wert von 0,002 angegeben.
- 6) Es wurde die Differenz der beobachteten mittleren Veränderungen zur Baseline berechnet. Für die Berechnung des 95%-KI wird seitens des pU ein gemischtes Modell verwendet, welches Terme für die Behandlung und die Studienvisite (bzw. gemäß externem SAP Zeit und Behandlung) einschließt. Konkrete Angaben zum statistischen Modell, beispielsweise zur Kovarianzstruktur, konnten nicht identifiziert werden.

Abkürzungen: ITT: Intention-to-Treat; Jahr 1B: Jahr 1 nach Behandlung des zweiten Auges in der Interventionsgruppe; Jahr 1C: Jahr 1 nach Baseline in der Kontrollgruppe; k.A.: keine Angaben; KI: Konfidenzintervall; MLMT: Mult-Luminanz-Mobilitätstest; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; SD: Standardabweichung.

Der pU gibt im Studienbericht an, dass von der ITT-Population in der Interventionsgruppe 13 der 21 Personen (61,9 %) zu Jahr 1 den MLMT bei Beleuchtungsstärke 1 Lux bestanden haben. In der Kontrollgruppe hat keine Person den Test bei 1 Lux bestanden. Bis auf eine Person, bei der der Wert gleich blieb, konnte bei allen Teilnehmenden in der Interventionsgruppe eine positive Änderung im bilateralen Mobilitätsscore zu Jahr 1 beobachtet werden. Bei 8 Personen (31,1 %) konnte ein Änderungswert von 1 festgestellt werden, wobei bei 4 dieser Personen (19,0 %) kein noch höherer Änderungswert möglich war. Bei 11 Personen (52,4 %) konnte ein Änderungswert > 1 beobachtet werden.

Bei 4 der 10 in die Studie randomisierten Personen in der Kontrollgruppe (40,0 %) konnte eine Verbesserung im Score beobachtet werden. Bei einer dieser Person trat eine Verschlechterung um mehr als einen Punkt im Mobilitätsscore zu Jahr 1 auf. Bei 2 Personen trat keine Änderung auf und bei 3 Personen trat eine Verschlechterung auf.

Sensitivitätsanalysen

Es wurden Sensitivitätsanalysen mit der mITT-/Sicherheitspopulation (Intervention: N = 20, Kontrolle: N = 9) und PP-Population (Intervention: N = 19, Kontrolle: N = 9) durchgeführt und es zeigten sich mit der ITT-Population konsistente Ergebnisse.

2.6.4 Lebensqualität

Es lagen kein validen und für die Nutzenbewertung relevanten Endpunkte in der Kategorie Lebensqualität vor.

2.6.5 Sicherheit

Die zusammenfassende Darstellung der UE anhand der Sicherheitspopulation ist in Tabelle 15 dargestellt. Eine separate Darstellung der UE in zeitliche Abschnitte (beispielsweise unmittelbar nach der Behandlung, ab Beginn der Studienvisiten bis zum Studienende) wurden nicht durchgeführt. Im Dossier wurden zusätzlich post hoc Effektschätzer für UE berechnet, wobei einige Angaben fehlen. Es wurden keine a priori definierten UE von besonderem Interesse angegeben.

UE wurden in der Kontrollgruppe bereits zu Baseline, in der Interventionsgruppe erst ab der ersten Injektion erfasst. Die durchschnittliche Wartezeit von Randomisierung bis zur ersten Injektion betrug durchschnittlich 34,3 Tage (SD: 27,9) bzw. im Median 22 Tage (min; max: 6; 78). Ähnliche Angaben zur Kontrollgruppe konnten nicht identifiziert werden. Konkrete Angaben zur Beobachtungszeit beider Behandlungsgruppen bis zur Visite Jahr 1 sowie der Studiendauer konnten nicht identifiziert werden.

Tabelle 15: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE zu Jahr 1 nach Baseline in der Studie 301 (Sicherheitspopulation)

Patienten mit mindestens einem ...	Intervention: Voretigen Neparvovec ¹⁾ (N = 20) n (%)	Kontrolle: Beobachtendes Abwarten ¹⁾ (N = 9) n (%)	RR [95%-KI] ²⁾ ; p-Wert ³⁾
UE	20 (100)	9 (100)	– ⁴⁾
UE ≥ Grad 3 ⁵⁾	11 (55,0)	2 (22,2)	2,48 [0,80; 24,37]; 0,130
Schwere UE ⁶⁾	3 (15,0)	0 (0,0)	k.A. ⁷⁾
SUE	2 (10,0)	0 (0,0)	k.A. ⁷⁾ ; 1,000
UE, die zum Studienabbruch führten	0 (0,0)	0 (0,0)	n.b.
UE, die zum Tod führten ⁸⁾	0 (0,0)	0 (0,0)	n.b.

¹⁾ Unterschiedlicher Beginn des Erhebungszeitraum zwischen beiden Gruppen: Die Erhebung begann in der Interventionsgruppe ab der ersten Injektion und in der Kontrollgruppe ab der Baseline-Untersuchung. Weitere Angaben zur gesamten Beobachtungs-/Studiendauer konnten nicht identifiziert werden.

²⁾ Vom pU post hoc berechneter Effektschätzer mittels Vier-Felder-Tafel. Das KI wurde mit einer exakten Methode (Tail-Methode) berechnet

³⁾ P-Wert berechnet post hoc anhand des exakten Tests nach Fischer.

⁴⁾ Patientenrelevanz unklar.

⁵⁾ Eine zusammenfassende Darstellung der UE ≥ Grad 3 wird im Dossier vorgenommen. Weitere Angaben zu den konkreten UE konnten nicht identifiziert werden.

⁶⁾ Es wurden nur schwere, nicht aber milde oder moderate UE dargestellt.

⁷⁾ Es wurden seitens des pU keine Effektschätzer berechnet.

⁸⁾ Da Todesfälle als UE erfasst wurden und nicht als separater Endpunkt definiert waren, werden sie unter Sicherheit dargestellt.

Abkürzungen: k.A.: keine Angaben, KI: Konfidenzintervall; n.b.: nicht berechnbar; pU: pharmazeutischer Unternehmer; RR: Relatives Risiko; (S)UE: (Schwerwiegende/s) Unerwünschte/s Ereignis/se.

Unerwünschte Ereignisse

Tabelle 16: UE mit Inzidenz ≥ 10 % der Sicherheitspopulation in der Studie 301 innerhalb eines Jahres nach Baseline

MedDRA-Systemorganklasse Preferred Term	Intervention: Voretigen Neparvovec (N = 20) ¹⁾ n (%)	Kontrolle: Beobachtendes Abwarten (N = 9) ¹⁾ n (%)	RR [95%-KI] ²⁾ p-Wert ³⁾
Störungen des Blut und lymphatischen Systems	9 (45,0)	0 (0,0)	k.A. ⁴⁾
Leukozytose	9 (45,0)	0 (0,0)	
Erkrankungen des Auges	10 (50,0)	1 (11,1)	4,50 [0,95; 124,40]; 0,096
Katarakt	3 (15,0)	0 (0,0)	
Entzündungen des Auges	2 (10,0)	0 (0,0)	
Photopsie	0 (0,0)	1 (11,1)	
Netzhautriss	2 (10,0)	0 (0,0)	
Erkrankungen des Gastrointestinaltrakts	12 (60,0)	3 (33,0)	1,80 [0,76; 9,21] 0,245
Oberbauchschmerzen	2 (10,0)	0 (0,0)	
Diarrhö	2 (10,0)	1 (11,1)	
Gastritis	0 (0,0)	1 (11,1)	

MedDRA-Systemorganklasse Preferred Term	Intervention: Voretigen Neparvovec (N = 20) ¹⁾ n (%)	Kontrolle: Beobachtendes Abwarten (N = 9) ¹⁾ n (%)	RR [95%-KI] ²⁾ p-Wert ³⁾
Refluxerkrankung	0 (0,0)	1 (11,1)	
Übelkeit	6 (30,0)	1 (11,1)	
Erbrechen	8 (40,0)	2 (22,2)	
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	10 (50,0)	1 (11,1)	4,50 [0,95; 124,40] 0,096
Reaktionen auf die Prüfmedikation	2 (10,)	0 (0,0)	
Fieber	7 (35,0)	1 (11,1)	
Erkrankungen des Immunsystems	0 (0,0)	1 (11,1)	0,00 [0,00; 6,46] 0,310
Saisonal bedingte Allergie	0 (0,0)	1 (11,1)	
Infektionen und parasitäre Erkrankungen	11 (55,0)	4 (44,4)	1,24 [0,57; 6,56] 0,700
Ohrinfektion	1 (5,0)	1 (11,1)	
Nasopharyngitis	7 (35,0)	2 (22,2)	
Infektion der oberen Atemwege	2 (10,0)	3 (33,3)	
Verletzung; Vergiftung und durch Eingriffe bedingte Komplikationen	5 (25,0)	2 (22,2)	1,13 [0,26; 7,34] 1,000
Tierbiss	2 (10,0)	0 (0,0)	
Exkoration	0 (0,0)	1 (11,1)	
Augenverletzung	0 (0,0)	1 (11,1)	
Untersuchungen	7 (35,0)	1 (11,0)	3,15 [0,60; 82,84] 0,3715
Erhöhter Augeninnendruck	4 (20,0)	0 (0,0)	
Gewichtsabnahme	0 (0,0)	1 (11,1)	
Stoffwechsel- und Ernährungsstörungen	2 (10,0)	0 (0,0)	k.A.
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	1 (5,0)	1 (11,1)	0,45 [0,01; 14,86] 0,532
Rückenschmerzen	0 (0,0)	1 (11,1)	
Nackenschmerzen	0 (0,0)	1 (11,1)	
Erkrankungen des Nervensystems	10 (50,0)	3 (33,3)	1,50 [0,60; 6,61] 0,453
Kopfschmerzen	7 (35,0)	2 (22,2)	
Psychiatrische Erkrankungen	2 (10,0)	1 (11,1)	0,90 [0,08; 24,37] 1,000
Angst	0 (0,0)	1 (11,1)	
Emetophobie	0 (0,0)	1 (11,1)	
Erkrankungen der Nieren und Harnwege	3 (15,0)	1 (11,1)	1,35 [0,15; 34,78] 1,000
Hämaturie	3 (15,0)	1 (11,1)	
Erkrankungen der Geschlechtsorgane und der Brustdrüse	3 (15,0)	0 (0,0)	k.A.

MedDRA-Systemorganklasse Preferred Term	Intervention: Voretigen Neparvovec (N = 20) ¹⁾ n (%)	Kontrolle: Beobachtendes Abwarten (N = 9) ¹⁾ n (%)	RR [95%-KI] ²⁾ p-Wert ³⁾
Erkrankungen der Atemwege, des Brustraums und Mediastinums	10 (50,0)	5 (55,6)	0,90 [0,43; 2,49] 1,000
Husten	6 (30,0)	1 (11,0)	
Nasenbluten	2 (10,0)	0 (0,0)	
Verstopfte Nase	2 (10,0)	0 (0,0)	
Oropharyngealer Schmerz	6 (30,0)	4 (44,0)	
Erkrankungen der Haut und des Unterhautzellgewebes	2 (10,0)	1 (11,1)	0,90 [0,08; 24,37] 1,000
Akne	0 (0,0)	1 (11,1)	
Ekzem	0 (0,0)	1 (11,1)	
Gefäßkrankungen	1 (5,0)	1 (11,1)	0,45 [0,01; 14,86] 0,532
Bluthochdruck	1 (5,0)	1 (11,1)	

¹⁾ Unterschiedlicher Beginn des Erhebungszeitraums zwischen beiden Gruppen: Die Erhebung begann in der Interventionsgruppe ab der ersten Injektion und in der Kontrollgruppe ab der Baseline-Untersuchung. Weitere Angaben zur gesamten Beobachtungs-/Studiendauer konnten nicht identifiziert werden.

²⁾ Post hoc berechneter Effektschätzer mittels Vier-Felder-Tafel. Das KI wurde mittels exakter Methode berechnet.

³⁾ P-Wert berechnet post hoc anhand des exakten Tests nach Fischer.

⁴⁾ Keine Berechnung der Effektschätzer seitens des pU.

Abkürzungen: k.A.: keine Angaben; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; pU: pharmazeutischer Unternehmer; RR: Relatives Risiko; UE: Unerwünschte/s Ereignis/se.

Schwerwiegende unerwünschte Ereignisse

Tabelle 17: SUE mit Inzidenz $\geq 5\%$ in der Studie 301 innerhalb eines Jahres nach Baseline (Sicherheitspopulation)

MedDRA-Systemorganklasse Preferred Term	Intervention: Voretigen Neparvovec ¹⁾ (N = 20) n (%)	Kontrolle: Beobachtendes Abwarten ¹⁾ (N = 9) n (%)	RR [95%-KI] ²⁾ p-Wert ³⁾
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	2 (10,0)	0 (0,0)	k.A.
Reaktionen auf das Arzneimittel	2 (10,0)	0 (0,0)	
Erkrankungen des Nervensystems	1 (5,0)	0 (0,0)	k.A.
Krampfanfall	1 (5,0)	0 (0,0)	

¹⁾ Unterschiedlicher Beginn des Erhebungszeitraum zwischen beiden Gruppen: Die Erhebung begann in der Interventionsgruppe ab der ersten Injektion und in der Kontrollgruppe ab der Baselineuntersuchung. Weitere Angaben zur gesamten Beobachtungs-/Studiendauer konnten nicht identifiziert werden.

²⁾ Post hoc berechneter Effektschätzer mittels Vier-Felder-Tafel. Das KI wurde mittels exakter Methode berechnet.

³⁾ P-Wert berechnet post hoc anhand des exakten Tests nach Fischer.

Abkürzungen: k.A.: keine Angaben; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; RR: Relatives Risiko; SUE: Schwerwiegende/s unerwünschte/s Ereignis/se.

Schwere unerwünschte Ereignisse

In Tabelle 18 sind schwere UE aus dem Studienbericht dargestellt. Milde und moderate UE wurden nicht dargestellt.

Tabelle 18: Schwere UE mit Inzidenz $\geq 5\%$ in der Studie 301 innerhalb eines Jahres nach Baseline (Sicherheitspopulation)

MedDRA-Systemorganklasse Preferred Term	Intervention: Voretigen Neparvovec ¹⁾ (N = 20) n (%)	Kontrolle: Beobachtendes Abwarten ¹⁾ (N = 9) n (%)	RR [95%-KI] ²⁾ p-Wert ²⁾
Erkrankungen des Herzens	1 (5,0)	0 (0,0)	k.A.
Tachykardie	1 (5,0)	0 (0,0)	
Erkrankungen des Gastrointestinaltrakts	2 (10,0)	0 (0,0)	k.A.
Übelkeit	2 (10,0)	0 (0,0)	
Erbrechen	2 (10,0)	0 (0,0)	
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	2 (10,0)	0 (0,0)	k.A.
Reaktionen auf das Arzneimittel	2 (10,0)	0 (0,0)	
Brustschmerzen	1 (5,0)	0 (0,0)	
Erkrankungen des Nervensystems	1 (5,0)	0 (0,0)	k.A.
Krampfanfall	1 (5,0)	0 (0,0)	

¹⁾ Unterschiedlicher Beginn des Erhebungszeitraum zwischen beiden Gruppen: Die Erhebung begann in der Interventionsgruppe ab der ersten Injektion und in der Kontrollgruppe ab der Baselineuntersuchung. Weitere Angaben zur gesamten Beobachtungs-/Studiedauer konnten nicht identifiziert werden.

²⁾ Keine Berechnung der Effektschätzer seitens des pU.

Abkürzungen: k.A.: keine Angaben; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; pU: pharmazeutischer Unternehmer; RR: Relatives Risiko; UE: Unerwünschte/s Ereignis/se.

Unerwünschte Ereignisse \geq Grad 3

Im Dossier Modul 4 wurden zusammenfassende Ergebnisse zu UE \geq Grad 3 dargestellt. Insgesamt trat bei 11 Personen in der Interventionsgruppe und 2 Personen im Kontrollarm ein UE vom Grad 3/4 auf. Woher diese Angaben stammen, kann nicht nachvollzogen werden. Unklar bleibt, inwieweit diese Angaben mit den oben dargestellten schweren UE bzw. mit der Schweregradeinteilung in Orientierung an der WHO (World Health Organization) zusammenhängen.

3 Methodische Anmerkungen und Bewertung der Unterlagen

3.1 Zulassungsstatus und Zulassungspopulation von Studie 301

Voretigen Neparvovec ist zugelassen zur Behandlung von erwachsenen und pädiatrischen Patientinnen und Patienten mit Sehverlust aufgrund einer erblichen Netzhautdystrophie, die auf nachgewiesenen biallelischen RPE65-Mutationen beruht, und die über ausreichend lebensfähige Netzhautzellen verfügen. Die zulassungsrelevante Dosis beträgt $1,5 \times 10^{11}$ Vg. Es wird sequentiell ein Volumen von je 0,3 ml ins Auge injiziert. Der Abstand zwischen den Verabreichungen in beiden Augen beträgt 12 ± 6 Tage; es handelt sich jeweils um eine einmalige Applikation.

An der Studie 301 konnten Personen ab einem Alter von mindestens 3 Jahren teilnehmen. Es wurden Personen zwischen 4 und 44 Jahren in die Studie eingeschlossen. Die Personen waren in beiden Behandlungsgruppen durchschnittlich 15 Jahre alt. Es wurden überwiegend Kaukasier (ca. 70 %) eingeschlossen. Die Teilnehmenden stammten zum Großteil aus dem nordamerikanischen Raum (86 % in der Interventionsgruppe und 60 % in der Kontrollgruppe), insgesamt nahmen 16,1 % Personen aus dem europäischen Raum an der Studie teil. Eine Übertragbarkeit auf den deutschen Versorgungsraum für Personen zwischen 4 und 44 Jahren erscheint plausibel. Unklar bleibt die Übertragbarkeit bei Patientinnen/Patienten, die jünger bzw. älter sind.

Als Ein-/Ausschlusskriterien wurden detaillierte ophthalmologische Kriterien vorgegeben, um laut pU sowohl eine ausreichend lebensfähige Netzhautfunktion sicherzustellen, aber auch um eine bestimmte Krankheitsschwere untersuchen zu können. Gemäß Modul 4 sollten Personen mit einer vollständig atrophierten Netzhaut, die von einer Therapie mit der Prüfmedikation nicht profitieren, ausgeschlossen werden, dabei aber Personen mit nur wenigen lebensfähigen Netzhautzellen trotzdem eine Therapie ermöglicht werden. Als Begründung für die Auswahl der Kriterien wird auf klinische Expertise verwiesen. Zusätzlich sollte der MLMT durchführbar sein. Dabei soll sichergestellt werden, dass Personen mit einer annähernd normal-funktionierenden Netzhaut (Bestehen bei 1 Lux) nicht teilnehmen können, aber auch Personen, bei denen keine ausreichende Netzhautfunktion vorhanden ist (nicht Bestehen bei 400 Lux) an der Studie nicht teilnehmen können. Es bleibt unklar, ob Personen, die zwar vom Anwendungsgebiet umfasst, aber nicht in der Lage waren den MLMT durchzuführen, an der Studie hätten teilnehmen können.

3.2 Design und Methodik der Studie

Bei der Studie 301 handelt es sich um eine randomisierte, multizentrische, offene Studie, bei der die Prüfmedikation im Vergleich zu beobachtendem Abwarten verglichen wurde. Die Randomisierung erfolgte im Verhältnis 2:1 in Form einer Block-Randomisierung (Blockgröße von 3), stratifiziert nach Alter der Patientinnen/Patienten (≥ 10 vs. < 10 Jahre) und Ergebnislevel des Mobilitätstests zum Screening mit dem am schlechtesten im Test abschneidenden Auge (Bestehen bei ≥ 125 vs. < 125 Lux). Der primäre Endpunkt war definiert als die Veränderung im MLMT bilateral gemessen nach Jahr 1B/C zwischen den Behandlungsgruppen. Die in der Kontrollgruppe teilnehmenden Personen konnten, sofern die Ein-/Ausschlusskriterien zu Jahr 1 erfüllt waren, die Prüfmedikation erhalten. Bis auf den primären Endpunkt (MLMT) fand keine Verblindung der Bewertung der Testergebnisse statt. Es wurden 21 Personen in die Interventions- und 10 Personen in die Kontrollgruppe eingeschlossen. Die Studie wurde von 20 Personen in der Interventionsgruppe und von 9 Personen zu Jahr 1 in der Kontrollgruppe beendet. Zwei Patientinnen/Patienten (jeweils eine/r in beiden Behandlungsarmen) schieden bereits vor Bekanntgabe der Behandlungszuteilung aus. Der

Großteil der Personen (ca. 90 %) hat die zulassungsrelevante Dosierung bzw. das Volumen des Prüfpräparats in der Interventionsgruppe erhalten.

Aufgrund des offenen Studiendesigns wird das Verzerrungspotential als hoch eingeschätzt. Desweiteren konnten Imbalancen in der Begleitmedikation zwischen den Behandlungsgruppen beobachtet werden. Der Anteil an Personen mit Schmerzmittel, Anticholinergika, Kortikosteroide + Antiinfektiva in Kombination, war in der Interventionsgruppe um 42,4–85,7 % höher als in der Kontrollgruppe.

Angaben zu den Beobachtungszeiträumen bzw. zur Studiendauer zwischen den Behandlungsgruppen konnten nicht identifiziert werden. Die Zeiträume zwischen Screening und Baseline sowie in der Behandlungsgruppe zwischen Baseline und erster Injektion konnte jeweils bis zu 90 Tage betragen. Der Zeitraum zwischen den Visiten durfte nicht mehr als 90 Tage betragen. Für die Interventionsgruppe betrug der durchschnittliche Zeitraum zwischen Randomisierung und erster Injektion 34,3 Tage (SD: 27,9), was insgesamt für eine kurze Zeitspanne im Vergleich zu den Protokollvorgaben spricht. Ähnliche Angaben für die Kontrollgruppe konnten in den Studienunterlagen nicht identifiziert werden.

Zudem finden sich Hinweise auf Unterschiede in der Studiendurchführung zwischen den Zentren. Es wurden mehr prozedurabhängige Abweichungen im Studienzentrum IOWA berichtet als im Studienzentrum CHOP. Auch war die durchschnittliche Wartezeit im Studienzentrum IOWA mit 44,7 Tagen (SD: 30,7) länger als beim Studienzentrum CHOP (23,8 Tage (SD: 21,4)) und der Anteil an Personen, die in die Kontrollgruppe randomisiert wurden im Vergleich zum CHOP geringer (IOWA: 20 % vs. CHOP: 80 %).

Zwar waren Kriterien zur Auswahl des ersten Auges, das das Prüfpräparat erhalten sollte, nicht a priori im Studienprotokoll definiert, gemäß Studienbericht entschied das Prüfpersonal in Abhängigkeit der Sehschärfe und/oder Patientenpräferenz: Das erste Auge wies bei allen Teilnehmenden die schlechtere Sehschärfe auf; ausgenommen eine Person, bei der zu Baseline keine Unterschiede zwischen den Augen bestanden.

3.3 Wirksamkeit

Mortalität

Todesfälle waren nicht als eigenständiger Endpunkt in der Studie 301 definiert, sondern wurden als UE während der Studie erfasst. Es traten während der Erfassung von UE keine Todesfälle auf. (Details zur Operationalisierung siehe Tabelle 8).

Multi-Luminanz-Mobilitätstest

Die Analyse der Veränderung im Mobilitätsscore zu Jahr 1B/C im Vergleich zu Baseline zwischen den Behandlungsgruppen war als primärer Endpunkt definiert. Zusätzlich wurden die mittleren Änderungen zu Jahr 1 im Vergleich zu Baseline dargestellt. Für die Berechnung des p-Wertes wurde ein Wilcoxon-Rangsummen-Test angewandt, während für die Differenz ein gemischtes Modell mit Behandlung und Studienvisite herangezogen wurde. Genaue Angaben zu dem Modell, wie beispielsweise die verwendete Kovarianzstruktur, konnten nicht identifiziert werden. Der pU gibt als MID (Minimal Important Difference) eine Änderung um einen Punkt an. Da diese nicht auf Grundlage von ankerbasierten Verfahren abgeleitet wurde, sondern auf klinischen Annahmen beruht, wird auf eine Darstellung der Ergebnisse verzichtet (siehe Operationalisierung Tabelle 5).

In der Validierungsstudie deuten die Ergebnisse auf eine geringe Messunsicherheit und gute Reliabilität hin. Es konnte zudem eine ausreichende Konstruktvalidität nachgewiesen werden. Der Test konnte zwischen gesunden Personen und Personen mit Seheinschränkungen unterscheiden.

Beim MLMT zeigt sich zu Jahr 1 ein statistisch signifikantes Ergebnis zugunsten der Interventionsgruppe ($p < 0,001$) anhand der ITT-Population (Interventionsgruppe: $N = 21$; Kontrollgruppe: $N = 10$). Dieses Ergebnis wird trotz geringer Fallzahl durch Sensitivitätsanalysen mittels mITT- (Interventionsgruppe: $N = 20$, Kontrollgruppe: $N = 9$) und PP-Population (Interventionsgruppe: $N = 19$; Kontrollgruppe: $N = 9$) bestätigt. Keine Person in der Interventionsgruppe hat den Test zu Jahr 1 schlechter abgeschnitten als zu Baseline.

Aufgrund des offenen Studiendesigns wird das Verzerrungspotential bei dem subjektiven Endpunkt als hoch angesehen. So kann die Kenntnis der Behandlungszuordnung die Durchführung des Tests durch den/die Patienten/Patientin beeinflusst haben. Positiv ist die verblindete Beurteilung des Tests, die qualitätssichernden Maßnahmen während der Studie und ein standardisiertes Vorgehen bei der Durchführung des Tests. Eine hohe Übereinstimmung in den Testbeurteilungen konnte beobachtet werden, allerdings wurden die Ergebnisse der Validierungsstudie und der Studie 301 zusammen ausgewertet. Eine separate Darstellung der Ergebnisse konnte nicht identifiziert werden.

Kritisch anzumerken ist die fehlende Validierung des Mobilitätsscores und des finalen Kriteriums „Bestanden / nicht Bestanden“. Dieses erfolgte allerdings indirekt für die Beleuchtungsstärken für die Kriterien Genauigkeit und Zeitwert; mit Ausnahme der später zu Beleuchtungsstärke 250 Lux zusammengefassten Stärken 200 und 250 Lux. Zu diesen Beleuchtungsstärken kann keine abschließende Beurteilung durchgeführt werden.

Ein ähnlicher Deckeneffekt zu niedrigen Beleuchtungsstärken bzw. eine ähnliche Verteilung der Teilnehmenden auf die Beleuchtungsstärken konnte sowohl in der Validierungsstudie als auch in der Studie 301 zu Baseline identifiziert werden. Der Deckeneffekt spiegelt sich gemäß Studienbericht anschließend auch zu Jahr 1 wider: 4 der 8 Personen (19,0 %) in der Interventionsgruppe mit einem Änderungswert von 1 in der Studie 301, konnten keinen höheren Änderungswert erzielen, da diese Personen bereits zu Baseline einen niedrigen Mobilitätswert aufwiesen.

Im Studienbericht wurde angemerkt, dass bei 4 der 10 Patientinnen/Patienten (40 %) der Kontrollgruppe ein positives Ergebnis beobachtet wurde, was durch Lerneffekte, verbesserte Fähigkeit sich zu Bewegen durch Reifeprozesse bzw. Variabilität in der Beurteilung beim finalen Kriterium „Bestanden/nicht bestanden“ erklärt wurde. Es ist nicht auszuschließen, dass diese Faktoren auch zu einer Verbesserung in der Interventionsgruppe beigetragen haben.

Der Anteil an Personen mit prozedurbedingten Abweichungen war gemäß Dossier in der Interventionsgruppe größer als in der Kontrollgruppe. Insgesamt kam es zu 5 Prozedurabweichungen in der Interventionsgruppe (3 zu Baseline und 2 zu Jahr 1) und zu einer in der Kontrollgruppe. Konkrete Angaben zur Anzahl an Personen konnten nicht identifiziert werden.

Zusammenfassende Einschätzung zur Wirksamkeit

Bei dem primären Endpunkt „Veränderung im MLMT-Score für beide Augen (bilateral) zu Jahr 1B/C zwischen den Behandlungsgruppen“ konnten statistisch signifikante Ergebnisse zugunsten der Interventionsgruppe festgestellt werden. Der pU gibt als MID eine Änderung um einen Punkt an. Da diese nicht auf Grundlage von ankerbasierten Verfahren abgeleitet wurde, sondern auf klinischen Annahmen beruht, wird auf eine Darstellung der Ergebnisse verzichtet. Das

Verzerrungspotential wird aufgrund des offenen Studiendesigns als hoch eingeschätzt. Positiv anzumerken sind aber die verblindete Bewertung, die durchgeführten qualitätssichernden Maßnahmen und das in der SOP beschriebene standardisierte Vorgehen. Unklar bleiben der erhöhte Anteil an prozedurbedingten Abweichungen in der Durchführung in der Interventionsgruppe. Kritisch anzumerken ist die fehlende Validierung des finalen Kriteriums „bestanden / nicht bestanden“, des Mobilitätsscores und der Beleuchtungsstufe 250 Lux sowie der nicht adäquaten Untersuchung der Änderungssensitivität bzw. MID. Zudem traten in der Validierungsstudie und der Studie 301 Deckeneffekt auf.

3.4 Lebensqualität

Es konnten keine bewertungsrelevanten Endpunkte in dieser Kategorie identifiziert werden.

3.5 Sicherheit

Bei allen Personen in beiden Behandlungsgruppen trat mindestens ein UE auf. Schwere UE, wie im Studienbericht dargestellt, und schwerwiegende UE (SUE) traten nur in der Interventionsgruppe bei 3 bzw 2 Personen, nicht aber in der Kontrollgruppe auf. Der pU hat post hoc für einzelne UE auf Ebene der Systemorganklassen sowie für die zusammengefassten UE \geq Grad 3 relative Risiken berechnet. Der Anteil an Personen mit UE auf Ebene der Systemorganklassen war in der Interventionsgruppe überwiegend größer als in der Kontrollgruppe, wobei die Unterschiede statistisch nicht signifikant waren. Eine abschließende Beurteilung der Aussagesicherheit in der Analyse, wie im nächsten Abschnitt erläutert, ist derzeit nicht abschließend möglich. Diese ist bereits aufgrund der kleinen Fallzahl eingeschränkt.

Das Verzerrungspotential wird aufgrund des offenen Studiendesigns und der fehlenden Verblindung der Studienteilnehmenden und des Studienpersonals als hoch angesehen. Ein weiterer Grund ist der ungleiche Beginn der UE-Erfassung zwischen den Behandlungsmethoden. In der Interventionsgruppe wurden UE, anders als in der Kontrollgruppe, nicht ab Baseline sondern erst ab der ersten Injektion erfasst, die durchschnittlich 34,3 Tage (SD: 27,9) nach der Randomisierung stattfand. Ähnliche Angaben konnten für die Kontrollgruppe nicht identifiziert werden. Der Zeitraum zwischen Screening und Baseline-Visite konnte bis zu 90 Tage betragen. Der spätere Zeitpunkt der UE-Erfassung in der Interventionsgruppe gegenüber der Kontrollgruppe führt zu einer Verzerrung der Ergebnisse zugunsten der Interventionsgruppe. Für beide Behandlungsgruppen konnten keine Angaben zur Beobachtungsdauer identifiziert werden, weshalb die Vergleichbarkeit der Beobachtungszeiten zwischen den Gruppen derzeit nicht möglich ist. Bei unterschiedlichen Beobachtungsdauern wären beispielsweise andere Methoden zur Berechnung von Effektschätzern, wie beispielsweise Überlebenszeitanalysen, die geeignetere Wahl.

In der Interventionsgruppe war der Anteil der Personen, die Schmerzmittel, Anticholinergika und Glukokortikoide in Kombination mit Antiinfektiva als Begleitmedikation einnahmen höher als in der Kontrollgruppe, was ebenfalls zu einer Verzerrung der Ergebnisse beitragen kann. Zwei UE (Leukozytose, Hyperglykämien) wurden in der Interventionsgruppe nur ab einem Schweregrad ≥ 3 während der Einnahmedauer von Glukokortikoiden als UE erfasst, wodurch es zu einer nicht vollständigen Erhebung aller UE während der Behandlung mit dem Prüfpräparat (18–30 Tage, abhängig vom Zeitpunkt der zweiten Injektion) kam. Da es nur zwei UE betrifft, wird dieser Einfluss als gering gewertet.

Angaben zu UE \geq Grad 3 wurden nur für die zusammengefassten UE im Dossier Modul 4 gemacht. Es bleibt unklar, inwieweit die zusammengefassten UE \geq Grad 3 mit der Schweregradeinteilung aus dem Studienbericht zusammenhängt, da sich diese jeweils unterscheiden. Zusätzlich wurden im Studienbericht, sofern möglich, eine Schweregradeinteilung orientierend an der WHO-Toxizitätsskala vorgenommen. Diese Ergebnisse konnten in den Studienunterlagen nicht identifiziert werden.

Es wurden keine UE von besonderem Interesse a priori für die Studie 301 definiert.

Zusammenfassende Einschätzung zur Sicherheit

SUE und schwere UE traten nur in der Interventionsgruppe auf und der Anteil an Personen mit UE \geq Grad 3 war in der Interventionsgruppe höher als in der Kontrollgruppe. Unklar bleibt, weshalb nicht für alle zusammengefassten UE und UE auf Ebene der Systemorganklassen Effektschätzer berechnet wurden. Die zusammengefassten Ergebnisse zur Schweregradeinteilung im Dossier unterscheidet sich von der im Studienbericht. Aufgrund der geringen Fallzahl ist die Aussage-sicherheit der Ergebnisse bereits eingeschränkt.

Das Verzerrungspotential wird für den Endpunkt als hoch angesehen; Grund hierfür ist das offene Studiendesign. Zusätzlich liegt eine Verzerrung aufgrund des im Vergleich zur Kontrollgruppe späteren Beginns der Erfassung von UE in der Interventionsgruppe zugunsten der Interventionsgruppe vor. Imbalancen bei den Begleitmedikationen tragen ebenfalls zur Verzerrung bei. Ein Vergleich der Beobachtungszeiten zwischen den Behandlungsgruppen ist nicht möglich, da keine Angaben hierzu identifiziert werden konnten.

4 Anforderung an eine qualitätsgesicherte Anwendung

Die Vorgaben der Fachinformation sind zu berücksichtigen.

Die Behandlung muss durch einen Netzhautchirurgen mit Erfahrung in der Durchführung von Makulaoperationen eingeleitet und durchgeführt werden.

5 Zusammenfassung der Bewertung

Voretigen Neparvovec ist zugelassen zur Behandlung von erwachsenen und pädiatrischen Patientinnen und Patienten mit Sehverlust aufgrund einer erblichen Netzhautdystrophie, die auf nachgewiesenen biallelischen RPE65-Mutationen beruht, und die über ausreichend lebensfähige Netzhautzellen verfügen. Die Nutzenbewertung von Voretigen Neparvovec basiert auf der zulassungsbegründenden Studie 301. Es handelt sich bei der Studie um eine randomisierte, multizentrische, offene Studie, bei der das Prüfpräparat im Vergleich zu beobachtendem Abwarten verglichen wurde. Die Randomisierung erfolgte im Verhältnis 2:1 in Form einer Block-Randomisierung. Der primäre Endpunkt „Änderungen im MLMT bilateral gemessen“ wurde zu Jahr 1 im Vergleich zu Baseline zwischen den Behandlungsgruppen erhoben. Die Personen der Kontrollgruppe konnten, sofern die Ein-/Ausschlussgründe erfüllt waren, nach Jahr 1 die Prüfmedikation erhalten.

Die Ergebnisse der Studie werden in der folgenden Tabelle unter Berücksichtigung des Verzerrungspotentials zusammengefasst. Es wird dargestellt, ob unter Voretigen Neparvovec ein statistisch signifikant positiver Effekt bei hohem oder unklarem (\uparrow) oder niedrigem ($\uparrow\uparrow$) Verzerrungspotential bzw. ein statistisch signifikant negativer Effekt bei hohem oder unklarem (\downarrow) oder niedrigem ($\downarrow\downarrow$) Verzerrungspotential bzw. kein Unterschied (\leftrightarrow) gezeigt werden konnte.

Tabelle 19: Zusammenfassende Darstellung der Ergebnisse der Studie 301

Endpunkt	Ergebnis		Effekt
	Intervention: Voretigen Neparvovec (N = 21)	Kontrolle: Beobachtendes Abwarten (N = 10)	
Mortalität			
Todesfälle, n (%) ¹⁾ Relatives Risiko [95%-KI] ²⁾ p-Wert ³⁾	0 (0,0) n.b. n.b.	0 (0,0) n.b. n.b.	n.b.
Morbidität: MLMT zu Jahr 1B/C⁴⁾			
Baseline n Mittelwert (SD) Median (min; max)	21 3,1 (1,7) k.A.	10 2,9 (1,6) k.A.	↑
Jahr 1 ⁵⁾ N ⁶⁾ Mittelwert (SD) Änderung zu Baseline: Mittelwert (SD) min; max	21 4,9 (2,0) 1,8 (1,1) 0; 4	10 3,2 (1,8) 0,2 (1,0) -1; 2	
Jahr 1 Exakter p-Wert ⁷⁾	0,001		
Jahr 1 Mittlere Änderung Intervention - Kontrolle zu Jahr 1 [95%-KI] ⁸⁾	1,6 [0,7; 2,4]		

Endpunkt	Ergebnis		Effekt
	Intervention: Voretigen Neparvovec (N = 21)	Kontrolle: Beobachtendes Abwarten (N = 10)	
Sicherheit⁹⁾10)11)			
UE \geq Grad 3 ¹²⁾ , n (%)	11 (55,0)	2 (22,2)	n.b. ¹³⁾
Relatives Risiko [95%-KI] ²⁾ ; p-Wert ³⁾	2,48 [0,80; 24,37]; 0,130		
Schwere UE ¹²⁾¹⁴⁾ , n (%)	3 (15)	0 (0,0)	n.b.
Relatives Risiko [95%-KI]; p-Wert	k.A.		
SUE ¹⁴⁾	2 (10)	0 (0,0)	n.b.
Relatives Risiko [95%-KI]; p-Wert	k.A.		

¹⁾ Todesfälle wurden als UE erfasst.

²⁾ Post hoc berechneter Effektschätzer. Das KI wurde mittels exakter Methode berechnet.

³⁾ P-Wert berechnet post hoc anhand des exakten Tests nach Fischer.

⁴⁾ Die Endpunktbewertung wurde verblindet durchgeführt.

⁵⁾ Primärer Endpunkt.

⁶⁾ Der Änderungswert für den MLMT zu Jahr 1B/C im Vergleich zu Baseline von Personen, die vor Applikation der Prüfmedikation ausgeschieden sind, wurden mit 0 imputiert.

⁷⁾ Es wurde ein Wilcoxon-Rangsummen-Test mit exaktem zweiseitigem p-Wert verwendet.

⁸⁾ Es wurde die Differenz der beobachteten mittleren Veränderungen zur Baseline berechnet. Für die Berechnung des 95%-KI wird seitens des pU ein gemischtes Modell verwendet, welches Terme für die Behandlung und die Studienvisite einschließt. Konkrete Angaben zu dem statistischen Modell, wie beispielsweise zur Kovarianzstruktur konnten nicht identifiziert werden.

⁹⁾ Die UE-Erfassung in der Interventionsgruppe erfolgte erst ab der ersten Injektion, die durchschnittlich 34,3 Tage (SD: 27,9) nach Baseline stattfand. Dies führt zu einem kürzeren Erfassungszeitraums in der Behandlungsgruppe gegenüber der Kontrollgruppe, bei der die Erfassung bereits ab Baseline geplant war. Ähnliche Angaben zum Zeitraum Screening-Baseline für die Kontrollgruppe konnten nicht identifiziert werden.

¹⁰⁾ Erfassung von bestimmten UE nur ab einem bestimmten Schweregrad im Interventionsarm: Sofern Leukozytose und Hyperglykämie während der Gabe von Glukokortikoiden (18–30 Tage, je nach Zeitpunkt der zweiten Injektion) auftraten, wurden diese nur als UE berichtet sofern sie als Grad 3 oder höher eingestuft wurden.

¹¹⁾ Es haben sowohl in der Interventions- als auch Kontrollgruppe Personen eine Begleitmedikation erhalten. Allerdings war der Anteil an Personen in der Interventionsgruppe, die Schmerzmittel einnahmen (Anilide (Intervention vs. Kontrolle: 81,0 vs. 0,0 %), Propionsäure-Derivate (52,4 vs. 10,0 %)), Anticholinergika (47,6 vs. 0,0 %), Kortikosteroide + Antiinfektiva in Kombination (85,7 vs. 0,0 %) höher als in der Kontrollgruppe.

¹²⁾ Ergebnisse für UE \geq Grad 3 wurden nur zusammenfassend im Dossier berichtet. Unklar bleibt, inwieweit diese mit der Schweregradeinteilung im Studienbericht zusammenhängen, da die zusammenfassenden Ergebnisse sich unterschieden.

¹³⁾ Angaben zu Studien- bzw. Beobachtungsdauer für beide Behandlungsgruppen konnten nicht identifiziert werden. Die Eignung bestimmter Auswerteverfahren der post hoc berechneten Effektschätzer (wie Relatives Risiko) bzw. inwieweit bei ungleichen Beobachtungszeiten geeigneter Verfahren wie Überlebenszeitanalysen angewandt werden können, kann nicht überprüft werden.

¹⁴⁾ Seitens des pU wurden keine Effektschätzer bzw. p-Werte berechnet.

Abkürzungen: Jahr 1B: Jahr 1 nach Behandlung des zweiten Auges in der Interventionsgruppe; Jahr 1C: Jahr 1 nach Baseline in der Kontrollgruppe; k.A.: keine Angaben; KI: Konfidenzintervall; MLMT: Multi-Luminanz-Mobilitätstest; n.b.: nicht bestimmbar, pU: pharmazeutischer Unternehmer; SD: Standardabweichung; (S)UE: (Schwerwiegende/s) Unerwünschte/s Ereignis/se.

Referenzen

1. **Beck RW, Moke PS, Turpin AH, Ferris FL, SanGiovanni JP, Johnson CA, et al.** A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol* 2003;135(2):194-205.
2. **Blackhurst DW, Maguire MG.** Reproducibility of refraction and visual acuity measurement under a standard protocol. The Macular Photocoagulation Study Group. *Retina* 1989;9(3):163-169.
3. **Chung DC, McCague S, Yu ZF, Thill S, DiStefano-Pappas J, Bennett J, et al.** Novel mobility test to assess functional vision in patients with inherited retinal dystrophies. *Clinical & experimental ophthalmology* 2018;46(3):247-259.
4. **Csaky KG, Richman EA, Ferris FL.** Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci* 2008;49(2):479-489.
5. **European Medicines Agency (EMA).** EU Regulatory Workshop – Ophthalmology – Summary and Report [online]. Amsterdam (NED): EMA; 2012. [Zugriff: 25.04.2019]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/09/WC500131815.pdf.
6. **European Medicines Agency (EMA).** Luxturna (International non-proprietary name voretigene neparvovec): European public assessment report EMEA/H/C/004451/0000 [online]. 20.09.2018. Amsterdam (NED): EMA. [Zugriff: 25.04.2019]. URL: https://www.ema.europa.eu/en/documents/assessment-report/luxturna-epar-public-assessment-report_en.pdf.
7. **Grossman DC, Curry SJ, Owens DK, Barry MJ, Davidson KW, Doubeni CA, et al.** Vision screening in children aged 6 months to 5 years: US Preventive Services Task Force recommendation statement. *Jama* 2017;318(9):836-844.
8. **Hamel C.** Retinitis pigmentosa [online]. In: Orphanet. 06.2014. [Zugriff: 10.05.2019]. URL: https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=791.
9. **Holladay JT.** Visual acuity measurements. *J Cataract Refract Surg* 2004;30(2):287-290.
10. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).** Allgemeine Methoden; Vers. 5.0 [online]. 10.07.2017. Köln (GER): IQWiG; 2017. [Zugriff: 01.07.2019]. URL: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.
11. **IQVIA.** Voretigene Neparvovec: Statistical Analysis Plan; Efficacy and Safety; Request for additional analyses in preparation of a benefit assessment of pharmaceuticals in accordance with the Act on Restructuring of the Market for Medicinal Products (AMNOG) in Germany [unveröffentlicht]. 08.01.2019.
12. **Lange C, Feltgen N, Junker B, Schulze-Bonsel K, Bach M.** Resolving the clinical acuity categories "hand motion" and "counting fingers" using the Freiburg Visual Acuity Test (FrACT). *Graefes Arch Clin Exp Ophthalmol* 2009;247(1):137-142.
13. **Leone JF, Gole GA, Mitchell P, Kifley A, Pai AS, Rose KA.** Visual acuity testability and comparability in Australian preschool children: the Sydney Paediatric Eye Disease Study. *Eye (Lond)* 2012;26(7):925-932.
14. **Lorenz B, Preising M.** Leber congenital amaurosis [online]. In: Orphanet. 07.2015. [Zugriff: 10.05.2019]. URL: https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=GB&Expert=65.

15. **Messias K, Jagle H, Saran R, Ruppert AD, Siqueira R, Jorge R, et al.** Psychophysically determined full-field stimulus thresholds (FST) in retinitis pigmentosa: relationships with electroretinography and visual field outcomes. *Doc Ophthalmol* 2013;127(2):123-129.
16. **Novartis Pharma.** Dossier zur Nutzenbewertung gemäß § 35a SGB V: Voretigen Neparvovec (Luxturna) [unveröffentlicht]. 08.04.2019.
17. **Novartis Pharma Gmb H.** Fachinformation Luxturna®. Dummy. Stand Monat/20192019.
18. **Patel DE, Cumberland PM, Walters BC, Russell-Eggitt I, Rahi JS, group Os.** Study of Optimal Perimetric Testing in Children (OPTIC): Feasibility, Reliability and Repeatability of Perimetry in Children. *PLoS One* 2015;10(6):e0130895.
19. **Patino CM, Varma R, Azen SP, Conti DV, Nichol MB, McKean-Cowdin R, et al.** The impact of change in visual field on health-related quality of life the los angeles latino eye study. *Ophthalmology* 2011;118(7):1310-1317.
20. **Roman AJ, Schwartz SB, Aleman TS, Cideciyan AV, Chico JD, Windsor EAM, et al.** Quantifying rod photoreceptor-mediated vision in retinal degenerations: dark-adapted thresholds as outcome measures. *Experimental eye research* 2005;80(2):259-272.
21. **Rosser DA, Cousens SN, Murdoch IE, Fitzke FW, Laidlaw DA.** How sensitive to clinical change are ETDRS logMAR visual acuity measurements? *Invest Ophthalmol Vis Sci* 2003;44(8):3278-3281.
22. **Shah N, Laidlaw DA, Rashid S, Hysi P.** Validation of printed and computerised crowded Kay picture logMAR tests against gold standard ETDRS acuity test chart measurements in adult and amblyopic paediatric subjects. *Eye (Lond)* 2012;26(4):593-600.
23. **Spark Therapeutics.** Mobility Test Validation Study: single-center observational study comparing the performance of normal-sighted and visually-impaired subjects on standardized obstacle courses; study report [unveröffentlicht]. 02.07.2015.
24. **Spark Therapeutics.** A Safety and Efficacy Study in Subjects with Leber Congenital Amaurosis (LCA) Using Adeno-Associated Viral Vector to Deliver the Gene for Human RPE65 to the Retinal Pigment Epithelium (RPE) [AAV2-hRPE65v2-301; statistical analysis plan; Version 4.0 [unveröffentlicht]. 19.08.2015.
25. **Spark Therapeutics.** A Safety and Efficacy Study in Subjects with Leber Congenital Amaurosis (LCA) Using Adeno-Associated Viral Vector to Deliver the Gene for Human RPE65 to the Retinal Pigment Epithelium (RPE) [AAV2-hRPE65v2-301]; clinical study report [unveröffentlicht]. 13.12.2016.
26. **Spark Therapeutics.** A Safety and Efficacy Study in Subjects with Leber Congenital Amaurosis (LCA) using Adeno-Associated Viral Vector to Deliver the Gene for Human RPE65 to the Retinal Pigment Epithelium (RPE) [AAV2-hRPE65v2-301]; clinical study report; appendix 16.1.1 protocol and protocol amendments [unveröffentlicht]. 13.12.2016.
27. **Talib M, Dagnelie G, Boon CJF.** Recording and Analysis of Goldmann Kinetic Visual Fields. In: Camiel JF, Boon JW (Eds). *Retinal Gene Therapy: Methods and Protocols*. New York: Springer; 2018. S. 327-338. (Methods in Molecular Biology: Band 1715).
28. **Told R, Baratsits M, Garhöfer G, Schmetterer L.** Early treatment diabetic retinopathy study (ETDRS) visual acuity. *Ophthalmologie* 2013;110(10):960-965.