

Nutzenbewertung

von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V

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

Wirkstoff: Ganaxolon

Datum der Veröffentlichung: 1. Oktober 2026

Version 1.0

Inhaltsverzeichnis

Tabellenverzeichnis	3
Abkürzungsverzeichnis	4
Hintergrund	6
1 Fragestellung	7
2 Auswahl und Methodik der relevanten Studien	8
2.1 Studienbasis für die Nutzenbewertung	8
2.2 Beschreibung der eingeschlossenen Studie	8
2.3 Endpunkte	14
2.3.1 Mortalität	15
2.3.2 Morbidität	15
2.3.3 Lebensqualität	24
2.3.4 Sicherheit	26
2.3.5 Erhebungszeitpunkte	27
2.4 Statistische Methoden	28
2.5 Verzerrungspotential auf Studien- und Endpunktebene	31
3 Ergebnisse der eingeschlossenen Studie	34
3.1 Studiencharakteristika und Studienmedikation	34
3.2 Mortalität	37
3.3 Morbidität	38
3.4 Lebensqualität	41
3.5 Sicherheit	43
4 Diskussion der Methodik und Ergebnisse	45
4.1 Design und Methodik der Studie	45
4.2 Zulassungsstatus und Zulassungspopulation	46
4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit	47
4.4 Mortalität	47
4.5 Morbidität	48
4.6 Lebensqualität	49
4.7 Sicherheit	49
5 Anforderung an eine qualitätsgesicherte Anwendung	50
6 Zusammenfassung der Nutzenbewertung	51
Referenzen	54

Tabellenverzeichnis

Tabelle 1:	Übersicht über die Studienbasis	8
Tabelle 2:	Charakterisierung der Studie Marigold	9
Tabelle 3:	Für die Nutzenbewertung relevante Protokolländerungen der Studie Marigold	12
Tabelle 4:	Charakterisierung der Intervention der Studie Marigold	13
Tabelle 5:	Zusammenfassung der Endpunktbewertung der Studie Marigold	14
Tabelle 6:	Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie Marigold	27
Tabelle 7:	Verzerrungspotential der Studie Marigold.....	31
Tabelle 8:	Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie Marigold.....	32
Tabelle 9:	Allgemeine Angaben; Studie Marigold.....	34
Tabelle 10:	Charakterisierung der Studienpopulation der Studie Marigold.....	34
Tabelle 11:	Begleitmedikation; Studie Marigold.....	37
Tabelle 12:	Veränderung der Häufigkeit jeglicher Anfälle der 17-wöchigen DBP gegenüber Baseline; Studie Marigold, ITT-Population (Datenschnitt: 19.05.2021)	38
Tabelle 13:	Responderanalysen zur Reduktion der Häufigkeit jeglicher Anfälle um $\geq 25\%$, $\geq 50\%$ und $\geq 75\%$ zu Woche 17; ITT-Population (Datenschnitt: 19.05.2021).....	39
Tabelle 14:	Veränderung der anfallsfreien Tage jeglicher Anfälle der 17-wöchigen DBP gegenüber Baseline; Studie Marigold, ITT-Population (Datenschnitt: 19.05.2021) ..	39
Tabelle 15:	Veränderung schwerer motorischer Anfälle der 17-wöchigen DBP gegenüber Baseline; Studie Marigold, ITT-Population (Datenschnitt: 19.05.2021) (ergänzend dargestellt)	40
Tabelle 16:	Anteil der Personen mit einer Verbesserung ¹ des globalen klinischen Betreuereindrucks der Verbesserung (CGI-I–Eltern/Betreuende) zu Woche 17 gegenüber Baseline (post hoc); ITT-Population (Datenschnitt: 19.05.2021)	41
Tabelle 17:	Lebensqualität gemäß dem QI-Disability-Fragebogen, Anzahl der Personen mit einer Verbesserung des QI-Disability um mindestens 15 % der Skalenspannweite zu Woche 17 (post hoc), ITT-Population (Datenschnitt: 19.05.2021)	41
Tabelle 18:	Ergebnisse der Nebenwirkungen – Zusammenfassung der UE unter Ausschluss krankheitsbezogener Ereignisse (Anfälle) während der 17-wöchigen DBP; Sicherheitspopulation, Studie Marigold (Datenschnitt: 19.05.2021)	43
Tabelle 19:	UE mit Inzidenz $\geq 10\%$ in der Studie Marigold; Sicherheitspopulation (Datenschnitt: 19.05.2021)	44
Tabelle 20:	SUE, die bei $\geq 5\%$ der Patientinnen und Patienten in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term in der Studie Marigold; Sicherheitspopulation (Datenschnitt: 19.05.2021).....	44
Tabelle 21:	Zusammenfassende Darstellung der Ergebnisse der Studie Marigold.....	51

Abkürzungsverzeichnis

ACTH	adrenocorticotropes Hormon
ADAMS	Anxiety, Depression and Mood Scale
Allo-S	Allopregnanolon-Sulfat
ALT	Alanin-Aminotransferase
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
AST	Aspartat-Aminotransferase
AVE	Average Variance Extracted
CBD	Cannabidiol
CDD	CDKL5 Deficiency Disorder (CDKL5-Mangelerkrankung)
CDKL5	Cyclin-dependent kinase-like 5
CGI-C	Caregiver Global Impression of Change
CGICA	Caregiver Global Impression of Change in Attention
CGI-CSID	Caregiver Global Impression of Change in Seizure Intensity/Duration
CGI-I	Clinical Global Impression of Improvement
CSHQ	Children's Sleep Habit Questionnaire
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
DBP	doppelblinde, placebokontrollierte Studienphase
DRF/SIF	Seizure Identification and Diagnostic Review Form
EEG	Elektroenzephalographie
EMA	European Medicines Agency
EPAR	European Public Assessment Report
ESC	Epilepsie-Konsortium
FI	Fachinformation
GBA	Gemeinsamer Bundesausschuss
GFR	geschätzte glomeruläre Filtrationsrate
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to Treat
IWRS	Interactive Web Response System
k. A.	keine Angabe
KI	Konfidenzintervall
MedDRA	Medical Dictionary for Regulatory Activities
N	Anzahl
OLE	offene einarmige Studienphase (Open-Label Extension)
PSI	Parenting Stress Index
PT	Preferred Term
pU	pharmazeutischer Unternehmer
QI-Disability	Quality of Life Inventory-Disability
RMSEA	Root Mean Square Error of Approximation
RR	Relatives Risiko

SAP	Statistischer Analyseplan
SD	Standardabweichung
SE	Standardfehler
SGB	Sozialgesetzbuch
SIF/DRF	Seizure Identification and Diagnostic Review Form
SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis
THC	Tetrahydrocannabinol
UE	Unerwünschtes Ereignis
ULN	oberer Grenzwert des Normalbereichs (Upper Limit of Normal)
USA	Vereinigte Staaten von Amerika
VerfO	Verfahrensordnung des GBA
VNS	Vagus-Nerv-Stimulation
ZNS	Zentralnervensystem

Hintergrund

Ganaxolon ist zugelassen als Arzneimittel zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. Gemäß § 35a Absatz 1 Satz 11 1. Halbs. SGB V gilt der medizinische Zusatznutzen durch die Zulassung als belegt.

Gemäß § 35a Absatz 2 SGB V entscheidet der Gemeinsame Bundesausschuss (GBA), ob er die Nutzenbewertung selbst durchführt oder das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) beauftragt. Ausgehend von der gesetzlichen Vorgabe in § 35a Absatz 1 Satz 11 1. Halbs. SGB V, dass der Zusatznutzen eines Orphan Drug durch die Zulassung als belegt gilt, hat der GBA in seiner Sitzung vom 15. März 2012 das Verfahren der Nutzenbewertung von Orphan Drugs dahingehend modifiziert, dass bei Orphan Drugs zunächst keine eigenständige Festlegung einer zweckmäßigen Vergleichstherapie mehr durch den GBA als Grundlage der insoweit allein rechtlich zulässigen Bewertung des Ausmaßes eines gesetzlich zu unterstellenden Zusatznutzens erfolgt. Vielmehr wird ausschließlich auf der Grundlage der Zulassungsstudien das Ausmaß des Zusatznutzens durch den GBA bewertet.

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

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

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

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

Die Nutzenbewertung wird am 01. Oktober 2026 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des GBA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der GBA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1 Fragestellung

Ganaxolon (Ztalmy®) ist ein Arzneimittel mit dem Status „Orphan Drug“. Gemäß § 35a Absatz 1 Satz 11 1. Halbs. SGB V gilt für Orphan Drugs der medizinische Zusatznutzen bereits durch die Zulassung als belegt. Der GBA bestimmt bei Orphan Drugs, die einen Umsatz von 30 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. Die Bewertung bezieht sich auf folgende Indikation [9]:

Zusatzbehandlung von epileptischen Anfällen im Zusammenhang mit einer Cyclin-abhängigen Kinase-ähnlichen 5 (CDKL5)-Mangelerkrankung (CDD) bei Patienten im Alter von 2 bis 17 Jahren. ZTALMY kann bei Patienten ab 18 Jahren weiter angewendet werden.

Nach anfänglicher gewichtsabhängiger Aufdosierung beträgt die zugelassene Maximaldosis (Erhaltungsdosis) gemäß Fachinformation für Patientinnen und Patienten ≤ 28 kg 63 mg/kg/Tag und für alle Personen > 28 kg maximal 1.800 mg/Tag.

2 Auswahl und Methodik der relevanten Studien

2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studienname (Nummer)	Studie mit Dossier vom pU eingereicht	Studie vom pU als relevant für die Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzenbewertung	Ausschlussgrund
Studien zum Wirkstoff				
Marigold (1042-CDD-3001) ¹⁾	Ja	Ja	Ja	-
1042-0900 ²⁾³⁾	Ja	Supportiv	Nein	unkontrollierte (Phase-IIa-)Proof-of-Concept-Studie bei Personen mit Enzephalopathien, inklusive CDD - kleine Patientenpopulation (n = 7; davon beendeten n = 4 die Studie) - Ein- und Ausschlusskriterien sowie erlaubte bzw. nicht erlaubte Begleitmedikationen abweichend zur pivotalen Studie

¹⁾ Zulassungsrelevante Studie gemäß EPAR.

²⁾ Im Rahmen des Zulassungsverfahrens bei der EMA ergänzend zu den Zulassungsstudien berücksichtigt.

³⁾ Der pivotalen Studie Marigold vorausgehende Proof-of-concept-Studie (Studiendauer der Open-Label-Phase: 08.05.2015 – 16.01.2018).

Abkürzungen: CDD: CDKL5-Mangel-Syndrom; EMA: European Medicines Agency; EPAR: European Public Assessment Report; pU: pharmazeutischer Unternehmer.

Die Studie 1042-0900 wird vor dem Hintergrund des Vorliegens der randomisiert kontrollierten Studie (RCT) und der geringen Aussagekraft der unkontrollierten Daten nicht für die Nutzenbewertung herangezogen.

Zur Nutzenbewertung für Ganaxolon herangezogene Unterlagen

- Dossier des pU zu Ganaxolon [10]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [7]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der Studie Marigold (1042-CDD-3001) [14, 15, 16]
- Fachinformation zu Ganaxolon [9]

2.2 Beschreibung der eingeschlossenen Studie

Die Nutzenbewertung von Ganaxolon im vorliegenden Anwendungsgebiet basiert auf der Zulassungsstudie Marigold (1042-CDD-3001). Die Studie und die Intervention werden in Tabelle 2 und Tabelle 4 charakterisiert.

Tabelle 2: Charakterisierung der Studie Marigold

Charakteristikum	Beschreibung
Design und Studienablauf	Design Multizentrische, randomisierte, doppelblinde, placebokontrollierte Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit von Ganaxolon bei Kindern und jungen Erwachsenen mit CDD. Randomisierung Unstratifizierte Randomisierung im Verhältnis 1:1 (Ganaxolon:Placebo) Studienablauf • Screening-Phase (14 Wochen)¹⁾ • Doppelblinde, placebokontrollierte Studienphase (DBP; 17 Wochen) ◦ Titrationsphase (4 Wochen) ◦ Erhaltungsphase (13 Wochen) • Sicherheits-Follow-up (2 Wochen) bei Studienabbruch²⁾ • Offene einarmige Studienphase (OLE; 216 Wochen) • Sicherheits-Follow-up (2 Wochen) bei Studienabbruch²⁾ Die 17-wöchige doppelblinde, placebokontrollierte Studienphase ist Gegenstand der vorliegenden Nutzenbewertung; die folgenden Angaben beziehen sich ausschließlich auf diese Studienphase.
Population	Wesentliche Einschlusskriterien • Patientinnen und Patienten zwischen 2–21 Jahren³⁾. • molekularanalytisch bestätigte (potenziell) pathogene CDKL5-Variante, früh einsetzende, schwer kontrollierbare Anfälle und neurologische Entwicklungsstörungen. • ≥ 16 schwere motorische Anfälle der primären Anfallstypen pro Monat in den 2 Monaten vor der Screening-Visite⁴⁾, darunter: ◦ bilateral tonischer Anfall (anhaltende motorische Aktivität ≥ 3 Sekunden); ◦ generalisierter tonisch-klonischer Anfall; ◦ atonischer Anfall/Sturzanfall; ◦ bilateral klonischer Anfall oder fokal zu bilateral tonisch-klonischer Anfall. • Versagen von ≥ 2 Antikonvulsiva zur Anfallskontrolle. • Stabiles Behandlungsschema mit (s. Tabelle 4): ◦ 0–4 anfallshemmenden Medikamenten⁵⁾ (inklusive mittelstarker oder starker antikonvulsiver Induktoren oder Inhibitoren wie Carbamazepin und Phenytoin etc.)⁶⁾, unverändert seit mind. 4 Wochen vor der Screening-Visite und ohne absehbare Dosisanpassung in DBP; ◦ Vagus-Nerv-Stimulation (VNS), ketogener Diät und modifizierter Atkins-Diät, unverändert seit 3 Monaten vor der Screening-Visite; ◦ Felbamat, in Verbindung mit gleichbleibender Leberfunktion (anhand AST bzw. ALT) und Hämatologie, unverändert seit ≥ 6 Monaten. • Patientinnen und Patienten mit chirurgisch implantierter VNS: ◦ Implantation der VNS ≥ 1 Jahr vor Screening-Visite; ◦ konstante Einstellungen seit 3 Monaten vor Screening-Visite und während DBP. Wesentliche Ausschlusskriterien • Frühere Exposition gegenüber Ganaxolon.

Charakteristikum	Beschreibung
	• Einnahme von: ◦ adrenocorticotropem Hormon (ACTH), Prednison oder anderen systemischen (nicht inhalativen/intranasalen oder topischen) Glucocorticoiden, < 28 Tagen vor der Screening-Visite; ◦ mäßigen oder starken Induktoren oder Inhibitoren (inklusive Johanniskraut) von CYP3A4/5/7, außer Antikonvulsiva mit mäßiger oder starker Induktion oder Inhibition, wenn diese mind. 4 Wochen stabil eingenommen werden (siehe Einschlusskriterien); ◦ Nahrungsergänzungsmitteln, wenn Einnahme instabil oder Beginn < 3 Monate vor der Screening-Visite. • Relevante Erkrankungen: ◦ West-Syndrom mit Hypsarrhythmie-Muster in EEG oder Anfälle vorwiegend vom Typ der Infantilen Spasmen; ◦ aktive Erkrankungen des ZNS (Infektionen, demyelinisierende Erkrankungen, degenerative neurologische Erkrankungen, progressive ZNS-Erkrankungen (durch bildgebende Verfahren bewertet)); ◦ Erkrankung oder Zustand (außer CDKL5), einhergehend mit einem erhöhten Risiko für den Patienten / die Patientin oder mit einer Beeinträchtigung des hämatologischen, kardiovaskulären, pulmonalen, renalen, gastrointestinalen oder hepatischen Systems bzw. anderen Zuständen, die die Absorption, die Verteilung, den Metabolismus oder die Ausscheidung des Prüfpräparats beeinträchtigen; ◦ Niereninsuffizienz: eGFR < 30 ml/min (mittels Cockcroft-Gault-Formel, pädiatrischen GFR-Rechners oder Bedside-Schwartz-Gleichung). • Nachweis von positivem THC oder CBD (mittels Urin- oder Plasmadrogenscreenings) am Tag der Screening-Visite und einem positiven Test auf THC oder CBD (mittels Plasmas) am Tag der Baseline-Visite ohne Verordnung von Epidiolex (oder gleichem Produkt unter anderem Namen), außer wenn Epidiolex in DBP stabil bis ≥ 1 Monat vor Screening und ohne absehbare Dosisanpassung. • AST oder ALT ◦ Studienbeginn: > 3 × des oberen Grenzwertes (ULN); ◦ während der Studie: bei Anstieg auf > 3 × ULN wöchentliche Kontrollen; bei fallender Tendenz Fortführung der Studie, andernfalls Ausschluss, wenn keine Senkung auf < 3 × ULN. • Gesamtbilirubinwerte ◦ Studienbeginn: oberhalb des ULN (Ausnahmen bei stabilem medizinischem Zustand durch medizinischen Monitor möglich); ◦ während der Studie: Gesamtbilirubinwerte > 1,5 × ULN. • Allopregnanolon-Sulfat-Plasmaspiegel (Allo-S) ≥ 6,0 ng/ml bei der Screening-Visite⁶.
Intervention und Zahl der Patientinnen und Patienten	Gescreent gesamt: N = 114 Randomisiert/eingeschlossen: N = 101 Intervention: N = 50 Kontrolle: N = 51
Ort und Zeitraum der Durchführung; Datenschnitte	Studienzentren 36 Studienzentren in 8 Ländern ⁷ : Australien, Frankreich, Israel, Italien, Polen, Russland, Vereinigtes Königreich, Vereinigte Staaten von Amerika (USA). Studienzeitraum • Erste Person, erste Visite (DBP): 30. Juni 2018 • Letzte Person, letzte Visite (OLE): 30. Juni 2023

Charakteristikum	Beschreibung
	Datenschnitte • Datenschnitt zum Ende der DBP: 31.07.2020 • Datenschnitt zum Ende der DBP⁸⁾ (Datenupdate)⁹⁾: 19.05.2021 • Erster Datenschnitt der OLE: 24.02.2021 • Zweiter Datenschnitt der OLE: 30.06.2022 • Finaler Datenschnitt der OLE: 01.09.2023 Es wird der präspezifizierte Datenschnitt zum Ende der DBP mit den aktualisierten Daten vom 19.05.2021 herangezogen. Die Studie ist zum Zeitpunkt der Nutzenbewertung abgeschlossen.
Endpunkte gemäß Studienprotokoll	Primärer Endpunkt Prozentuale Veränderung schwerer¹⁰⁾ motorischer Anfälle der primären Anfallsarten von Baseline zu Woche 17. Sekundäre Endpunkte (evaluiert von Baseline zu Woche 17) • Reduktion schwerer motorischer Anfälle um $\geq 50\%$; • Veränderung in CGI-I, CGI-CSID, CGI-C-Zielverhalten, CGICA; • Veränderung in anfallsfreien Tagen schwerer motorischer Anfälle; • Veränderung des längsten anfallsfreien Intervalls schwerer motorischer Anfälle. Explorative Endpunkte • Prozentuale Veränderung sekundärer¹¹⁾, tertiärer¹²⁾ und jeglicher¹³⁾ Anfälle sowie der einzelnen Anfallssubtypen; • Reduktion schwerer Anfälle um $\geq 25\%$ und $\geq 75\%$; • Reduktion schwerer Anfälle um $\geq 25\%$ und $\geq 75\%$; • Veränderung in anfallsfreien Tagen jeglicher Anfälle; • Veränderung des längsten anfallsfreien Intervalls schwerer motorischer Anfälle und jeglicher Anfälle; • CSHQ; • ADAMS; • Lebensqualität (QI-Disability, PSI¹⁴⁾); • Anteil an Personen mit Notfall-Medikation und Anzahl der Dosen; • EEG-Befunde. Sicherheit und Verträglichkeit • Unerwünschte Ereignisse.

¹⁾ Bestehend aus einer 8-wöchigen historischen (ab Woche -14) und einer anschließenden 6-wöchigen prospektiven Erhebung (ab Woche -6) mittels Anfallstagebuch.

²⁾ Personen, die die Studie entweder vorzeitig abgebrochen oder beendet haben und nicht an der Open-Label-Extensionsstudie teilnehmen, durchliefen eine 2-wöchige Taper-Periode. Während dieser Phase wurde die Dosis alle 3 Tage um 15 mg/kg/Tag (≤ 28 kg) bzw. 450 mg/Tag (> 28 kg) reduziert, bis die Studienmedikation vollständig abgesetzt war.

³⁾ Abweichung von der Zulassungspopulation gemäß FI. Eingeschlossen werden demnach auch Personen, welche älter als 17 Jahre sind. Eine Einnahme im Erwachsenenalter war gemäß FI jedoch u. U. möglich.

⁴⁾ Die Mindestanfallshäufigkeit galt entsprechend nicht für die 6-wöchige Baseline-Phase, welche als Referenzperiode zur Erfassung der Veränderungen dient, sondern ausschließlich für die insgesamt 8-wöchige historische Baseline-Phase.

⁵⁾ Abweichung von der Zulassungspopulation gemäß FI, welche eine Anwendung von Ganaxolon als Zusatztherapie zu anderen Antiepileptika vorsieht. Gemäß den beschriebenen Einschlusskriterien können auch Patientinnen und Patienten in die Studie eingeschlossen werden, welche zu Baseline keine (stabile) Behandlung mit antikonvulsiven/antiepileptischen Medikamenten erhalten.

⁶⁾ Eingeführt mit Protokoll-Amendment 1 vom 05.05.2019, etwa 10 Monate nach Studienbeginn (s. Tabelle 3)

- ⁷⁾ Die Anzahl der Studienzentren der jeweiligen Länder ist nicht berichtet.
- ⁸⁾ Ein weiterer Datenschnitt zum Datenbankschluss vom 01.09.2020 wurde durchgeführt. Es wurden jedoch keine Ergebnisse zu diesem Datenschnitt vorgelegt.
- ⁹⁾ Präspezifizierte primäre Wirksamkeitsanalyse mit Ende der 17-wöchigen doppelt verblindeten, placebokontrollierten Studienphase.
- ¹⁰⁾ Schwere – auch primär genannte – motorische Anfälle beinhalten beidseitig tonische oder klonische Anfälle, generalisierte tonisch-klonische Anfälle, atonische Anfälle/Sturzanfälle, fokal zu beidseitig tonisch-klonische Anfälle.
- ¹¹⁾ Sekundäre Anfälle sind definiert als fokale motorische Anfälle mit intaktem oder verändertem Bewusstsein oder fokale nicht-motorische Anfälle mit verändertem Bewusstsein.
- ¹²⁾ Tertiäre Anfälle sind definiert als fokale nicht-motorische Anfälle mit intaktem Bewusstsein, Absencen, myoklonische Anfälle oder epileptische Spasmen.
- ¹³⁾ Jegliche Anfälle umfassen sämtliche aufgetretene Anfallsarten.
- ¹⁴⁾ Der PSI wird als nicht patientenrelevant angesehen, da er die Belastung der betreuenden Eltern erfasst und nicht die der betroffenen CDD-Patientinnen und -Patienten.

Abkürzungen: ACTH: adrenocorticotropes Hormon; ADAMS: Anxiety, Depression and Mood Scale; Allo-S: Allopregnanolon-Sulfat; ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; CBD: Cannabidiol; CDD: CDKL5-Mangel-Syndrom; CDKL5: Cyclin-dependent kinase-like 5; CGI-C: Caregiver Global Impression of Change; CGI-CSID: Caregiver Global Impression of Change in Seizure Intensity/Duration; CGI-I: Clinical Global Impression of Improvement; CGICA: Caregiver Global Impression of Change in Attention; CSHQ: Children's Sleep Habit Questionnaire; DBP: doppelblinde, placebokontrollierte Studienphase; EEG: Elektroenzephalographie; eGFR: geschätzte glomeruläre Filtrationsrate; FI: Fachinformation; OLE: offene einarmige Studienphase (Open-Label Extension); PSI: Parenting Stress Index; QI-Disability: Quality of Life Inventory-Disability; THC: Tetrahydrocannabinol; ULN: oberer Grenzwert des Normalbereichs (Upper Limit of Normal); VNS: Vagus-Nerv-Stimulation; ZNS: Zentralnervensystem.

Protokolländerungen

Es wurden insgesamt 2 Änderungen des Originalprotokolls vom 08.06.2018 vorgenommen. Beide Protokoll-Amendments fanden zeitlich nach Einschluss der/des ersten Studienteilnehmenden statt (30.06.2018). Wie viele Patientinnen und Patienten zu diesem Zeitpunkt bereits in die Studie eingeschlossen waren, ist nicht bekannt. Wesentliche Änderungen des Amendments 1 vom 05.05.2019 sind Tabelle 3 zu entnehmen.

Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie Marigold

Amendment	Wesentliche Änderungen
Version 2 vom 26.05.2020, (Anzahl der bis dahin eingeschlossenen Personen: k. A.)	Keine wesentlichen Änderungen
Version 1 vom 05.05.2019, (Anzahl der bis dahin eingeschlossenen Personen: k. A.)	Datenschnitte - Interimsanalyse für den primären Endpunkt wurde hinzugefügt, wenn 50 bzw. 75 Patientinnen und Patienten die DBP beendet haben. Begleitmedikationen - Antikonvulsiva mit moderater oder starker induzierender bzw. hemmender Wirkung (z. B. Carbamazepin, Phenytoin) wurden als erlaubte Begleitmedikation hinzugefügt - Epidiolex (oder gleiches Produkt unter anderem Namen) wurde als erlaubte Begleitmedikation hinzugefügt.

Amendment	Wesentliche Änderungen
	Endpunkte • Aufnahme des sekundären Wirksamkeitsendpunkts („Anteil der Teilnehmenden mit ≥ 50 % Reduktion schwerer motorischer Anfälle“) • Hinzufügung von Subgruppenanalysen (Allo-S-Konzentration, Geschlecht) für den primären Endpunkt Statistik • Erhöhung der Fallzahl von 70 auf 100 Patientinnen und Patienten in Verbindung mit Anpassung der Power von 80 % auf 92 % und des Grenzwertes für statistisch signifikanten Effekt von 21 % auf 17,5 %.

Abkürzungen: Allo-S: Allopregnanolon-Sulfat; DBP: doppelblinde, placebokontrollierte Studienphase; k. A.: keine Angabe.

Charakterisierung der Intervention

Tabelle 4: Charakterisierung der Intervention der Studie Marigold

Ganaxolon	Placebo
Doppelblinde, randomisiert kontrollierte Behandlungsphase (17 Wochen)	
Ganaxolon 3 x täglich als orale Lösung. <u>Titrationsperiode, 4 Wochen (Tagesdosis)</u> • Woche 1 (Tag 1–7) ◦ ≤ 28 kg: 18 mg/kg ◦ > 28 kg: 450 mg • Woche 2 (Tag 8–14) ◦ ≤ 28 kg: 33 mg/kg ◦ > 28 kg: 900 mg • Woche 3 (Tag 15–21) ◦ ≤ 28 kg: 48 mg/kg ◦ > 28 kg: 1.350 mg • Woche 4 (Tag 22–28) ◦ ≤ 28 kg: 63 mg/kg ◦ > 28 kg: 1.800 mg <u>Erhaltungsperiode, 13 Wochen (Tagesdosis)</u> • ≤ 28 kg: 63 mg/kg • > 28 kg: 1.800 mg	Placebo • Wie bei Intervention, ohne den aktiven Wirkstoff.
Nicht erlaubte Begleitmedikation • systemische Kortikosteroide (ACTH, Prednison etc.) • CYP3A4/5/7-Modulatoren (Ketoconazol) • Cannabinoide¹⁾ Erlaubte Begleitmedikation • Antikonvulsiva (einschließlich aber nicht beschränkt auf Carbamazepin, Phenytoin, Felbamat)²⁾ • Rescue-Therapie³⁾ • topische bzw. intranasale Kortikosteroide zur Therapie von Hautreaktionen bzw. allergischer Rhinitis • Epidiolex (Cannabidiol; oder anderer Handelsname außerhalb der Vereinigten Staaten)	

¹⁾ Erlaubte Ausnahmen (siehe erlaubte Begleitmedikationen: Epidiolex oder äquivalentes Produkt außerhalb der USA, bei stabiler Behandlung seit mind. 4 Wochen)

²⁾ Stabile Behandlung seit mind. 4 Wochen vor dem Screening und ohne absehbare Dosisanpassung in DBP; Einnahme von Felbamat stabil seit > 6 Monaten sowie unveränderte Leberfunktion (AST, ALT) und Hämatologie

³⁾ Eine Definition bezüglich der Rescue-Therapie (z. B. Art und Dosierung der zugelassenen Medikamente) wird nicht berichtet.

Abkürzungen: ACTH: adrenocorticotropes Hormon; AST: Aspartat-Aminotransferase; ALT: Alanin-Aminotransferase; DBP: doppelblinde, placebokontrollierte Studienphase.

2.3 Endpunkte

In diesem Kapitel wird die Eignung der Endpunkte hinsichtlich Operationalisierung, Patientenrelevanz und Validität beurteilt. Dazu wurden das Herstellerdossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt. Alle Endpunkte der eingeschlossenen Studie (siehe Tabelle 2) wurden einer Prüfung unterzogen. Endpunkte, die in der nachfolgenden Tabelle 5 nicht gelistet sind, wurden weder vom pU noch im Rahmen der Nutzenbewertung als patientenrelevant bzw. bewertungsrelevant eingestuft. Tabelle 5 stellt das Ergebnis dieser Bewertung zusammenfassend dar.

Tabelle 5: Zusammenfassung der Endpunktbewertung der Studie Marigold

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Todesfälle ¹⁾	Mortalität	ja	ja
Epileptische Anfälle ²⁾	Morbidität	ja	ja
Globaler Betreuerereindruck (CGI)			
• CGI-CSID		ja	nein
• CGICA		ja	nein
• CGI-C-Zielverhalten		ja	nein
• CGI-I		ja	ja
CSHQ		ja	nein
ADAMS		ja	nein
QI-Disability	Lebensqualität	ja	ja
Unerwünschte Ereignisse	Sicherheit	ja	ja

¹⁾ Todesfälle wurden im Rahmen der Sicherheitserhebung erfasst.

²⁾ Verschiedene Operationalisierungen vorliegend. Primärer Endpunkt: Prozentuale Veränderung in der Häufigkeit der schweren motorischen Anfälle (primäre Anfallsarten). Der primäre Endpunkt wird ergänzend dargestellt.

Abkürzungen: ADAMS: Anxiety, Depression and Mood Scale; CGI-C: Caregiver Global Impression of Change; CGI-CSID: Caregiver Global Impression of Change in Seizure Intensity/Duration; CGI-I: Clinical Global Impression of Improvement; CGICA: Caregiver Global Impression of Change in Attention; CSHQ: Children's Sleep Habit Questionnaire; pU: pharmazeutischer Unternehmer; QI-Disability: Quality of Life Inventory-Disability.

2.3.1 Mortalität

Todesfälle

Der Endpunkt wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Die während der Studie Marigold aufgetretenen Todesfälle jeglicher Art wurden im Rahmen der Sicherheitserhebung vom Zeitpunkt der ersten Dosis bis 2 Wochen nach Ende der Taper-Phase in der 17-wöchigen doppelblinden Behandlungsphase als tödliche unerwünschte Ereignisse (UE) im Rahmen der Treatment-Emergent-Adverse-Events (TEAE) kontinuierlich erfasst.

Bewertung

Die Operationalisierung des Endpunktes ist nachvollziehbar.

Patientenrelevanz

Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen. Der Endpunkt wird als valide erachtet.

2.3.2 Morbidität

Epileptische Anfälle

Der Endpunkt „Epileptische Anfälle“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Epileptische Anfälle werden in der Studie Marigold hinsichtlich Häufigkeit, Art und der Klassifikation, ob einzelne oder gruppierte (mit zählbaren bzw. nicht zählbaren Anfällen) Anfälle vorliegen, unterschieden. Dabei werden die Anfallsarten studienspezifisch entsprechend den nachfolgend beschriebenen Anfallskategorien und den jeweils zugordneten Anfallsarten unterteilt:

- Schwere motorische (primäre) Anfälle:
 - o bilateral tonische Anfälle (anhaltende motorische Aktivität ≥ 3 Sekunden);
 - o generalisierte tonisch-klonische Anfälle;
 - o bilateral klonische Anfälle;
 - o atonische Anfälle/Sturzanfälle;
 - o fokal zu bilateral tonisch-klonische Anfälle.
- Sekundäre Anfälle:
 - o fokale motorische Anfälle mit intaktem oder verändertem Bewusstsein;
 - o fokale nicht-motorische Anfälle mit verändertem Bewusstsein.
- Tertiäre Anfälle:
 - o fokale nicht-motorische Anfälle mit intaktem Bewusstsein;
 - o Absenzen;
 - o myoklonische Anfälle;
 - o epileptische Spasmen.

Zusätzlich wird die Gesamtanfallsrate als Summe aller oben aufgeführten Anfälle erhoben, wobei Anfälle in einer Gruppe mit nicht zählbaren Anfällen als ein einzelner Anfall gezählt werden.

Die Anfälle werden täglich von der Betreuungsperson oder den Eltern der betroffenen Patientinnen und Patienten in einem elektronischen Tagebuch (eDiary) oder alternativ mit einem analogen Anfallskalender erhoben. Der Erhebungszeitraum umfasst die 14-wöchige Screening-Phase, bestehend aus der 8-wöchigen historischen und der 6-wöchigen prospektiven Baseline-Phase bis zum Ende der 17-wöchigen doppelblinden, placebokontrollierten Studienphase (DBP). Lediglich die Anfallsdaten der 6-wöchigen Baseline-Phase dienen jedoch als Referenzperiode zur späteren Messung der Veränderungen. Während der Studienvisiten sowie telefonisch wurden die Einträge im eDiary durch klinisches Prüfpersonal auf Richtigkeit sowie Vollständigkeit überprüft. Bei unklaren oder widersprüchlichen Einträgen werden Rückfragen an die entsprechende Betreuungsperson gestellt. Doppelte Einträge bleiben unberücksichtigt und fehlende Tagebucheinträge werden (außer im Rahmen von Sensitivitätsanalysen) nicht ersetzt (siehe Kapitel 2.4). Laut SAP unterbrechen Tage mit fehlenden Anfallsdaten das längste anfallsfreie Intervall nicht, werden hierbei jedoch nicht mitgezählt und werden ignoriert.

Um eine standardisierte Erfassung der Anfallsarten zu gewährleisten, mussten die Eltern oder die Betreuungspersonen ein verpflichtendes Dokument zur Anfallsidentifizierung und diagnostischen Überprüfung (Seizure Identification and Diagnostic Review Form, SIF/DRF) ausfüllen und dem Studienpersonal zur Verfügung stellen. Ohne dieses Dokument war keine Randomisierung möglich. Darüber hinaus sollten optional Videos der schweren motorischen Anfallsarten der Patientinnen und Patienten aufgenommen und ebenfalls bereitgestellt werden. Sowohl das SIF/DRF als auch die Videos (wenn verfügbar) wurden anschließend durch ein unabhängiges Epilepsie-Konsortium (ESC) evaluiert. Weitere Informationen, z. B. zur Zusammensetzung und anhand welcher Kriterien diese Überprüfung des Epilepsie-Konsortiums stattfand, gehen aus den Studienunterlagen nicht hervor.

Der pU stellt verschiedene Operationalisierungen zur Verfügung. Als primärer Endpunkt war definiert:

- Prozentuale Veränderung der Anfallshäufigkeit schwerer motorischer Anfälle pro 28 Tage gegenüber Baseline während der 17-wöchigen DBP.

Weitere Operationalisierungen während der 17-wöchigen DBP waren:

- Prozentuale Veränderung jeglicher Anfälle pro 28 Tage gegenüber Baseline.
- Prozentuale Veränderung der Anfallshäufigkeit sekundärer und tertiärer Anfälle pro 28 Tage.
- Prozentuale Veränderung der Anfallshäufigkeit der Anfallssubtypen (siehe oben) pro 28 Tage.
- Prozentuale Veränderung der anfallsfreien Tage:
 - schwerer motorischer Anfälle pro 28 Tage;
 - von Anfällen ohne schwere motorische Anfälle pro 28 Tage (post hoc);
 - jeglicher Anfälle pro 28 Tage.
- Veränderung des längsten konsekutiven anfallsfreien Intervalls:
 - schwerer motorischer Anfälle pro 28 Tage;
 - von Anfällen ohne schwere motorische Anfälle pro 28 Tage (post hoc);
 - jeglicher Anfälle pro 28 Tage.
- Responderanalysen ($\geq 25\%$, $\geq 50\%$, $\geq 75\%$) zur Anzahl der Patientinnen und Patienten mit einer Reduktion:
 - schwerer motorischer Anfälle pro 28 Tage;
 - jeglicher Anfälle pro 28 Tage.

Bewertung

Die Operationalisierung ist eingeschränkt nachvollziehbar.

Der pU nimmt entgegen der gängigen Einteilung epileptischer Anfälle (z. B. tonisch-klonische, konvulsive, nicht-konvulsive etc.) eine studienspezifische Schweregradeinteilung (schwer motorisch = primär, sekundär, tertiär) vor, begründet diese vorgenommen Einteilung jedoch nicht. Anhand welcher Klassifikation die Erfassung der Anfallsarten innerhalb des Anfallstagebuches umgesetzt werden sollte, geht aus den Studienunterlagen nicht eindeutig hervor. Es ist nicht gänzlich nachvollziehbar, ob die Betreuenden vor Studienbeginn eine Schulung zur Einschätzung der Anfallstypen erhalten haben und ob die Erfassung in den Anfallstagebüchern in Form der studieneigenen Anfallskategorien oder nach Anfallsart erfolgen sollte. Allerdings sollte vor Studienbeginn ein Dokument zur Anfallsidentifizierung und diagnostischen Überprüfung ausgefüllt werden, womit zumindest geprüft wurde, ob die Betreuungspersonen in der Lage waren, die Anfallstypen korrekt einzuschätzen. Aufbau und Inhalt des Dokumentes hingegen sind unklar. Zudem erfolgte eine Qualitätskontrolle der Tagebucheinträge durch das Prüfpersonal (per Telefon) und das unabhängige ESC. Anhand welcher Kriterien die Überprüfung durch das Prüfpersonal und das Konsortium erfolgte und ob die Tagebucheinträge immer durch dieselbe Person zu erfolgen hatten, ist nicht beschrieben. Darüber hinaus wurde ein anfallsfreies Intervall durch fehlende Werte nicht beendet, sondern fortgeführt, was zu einer Überschätzung der anfallsfreien Tage führen kann. Demgegenüber kann die Erfassung der Gesamtanfallsrate bei gruppierten und nicht zählbaren Anfällen als ein Anfall zu einer Untererfassung des tatsächlichen Anfallsgeschehens führen. Es ist zudem unklar, inwiefern die Kriterien zur Anfallserfassung des eDiarys mit dem alternativ eingesetzten analogen Anfallskalender übereinstimmten und somit vergleichbar waren.

Patientenrelevanz

Die CDD ist gekennzeichnet durch multiple Anfallstypen und eine hohe Anfallshäufigkeit. Dementsprechend ist eine Reduktion der Anfallshäufigkeit patientenrelevant. Das schließt jegliche Anfallsarten ein. Für die vorliegende Nutzenbewertung werden sämtliche erfassten Anfälle (z. B. schwere motorische, sekundäre, tertiäre Anfälle) als patientenrelevant angesehen. Da die schweren, sekundären und tertiären Anfälle von den jeglichen Anfällen umfasst werden, werden ausschließlich die jeglichen Anfälle berücksichtigt.

Validität

Die Art, die Klassifikation in einzeln oder gruppiert (sowie zählbar und nicht zählbar) sowie die Häufigkeit der Anfälle wurden mittels eines elektronischen oder analogen Tagebuches täglich erhoben. Unter Berücksichtigung des verpflichtenden SIF/DRF-Dokumentes zur Anfallsidentifizierung und der Überprüfung durch das Prüfpersonal sowie durch das ESC, erscheint die Erhebung insgesamt valide. Einschränkend zu berücksichtigen sind jedoch die fehlenden Informationen zum Inhalt des SIF/DRF sowie zur Zusammensetzung und konkreten Vorgehensweise des Epilepsie-Konsortiums, die Unklarheit inwiefern eine Schulung der Betreuungspersonen zur Anfallsidentifikation erfolgte sowie unklare Details zur praktischen Nutzung des Tagebuches, insbesondere hinsichtlich der Klassifikation in studieneigene Kategorien (siehe Operationalisierung). Grundsätzlich ist jedoch zu erwarten, dass die betreuenden Personen erfahren bzw. geschult in der Detektion und Klassifizierung von epileptischen Anfällen sind, sodass die Erhebung des Endpunktes als ausreichend valide eingeschätzt wird. Der Endpunkt „jegliche Anfälle“ umfasst entsprechend alle Anfallsarten, so dass eine potenziell mögliche fehlerhafte Zuordnung der Anfallsarten nicht zu einer Fehlerfassung führt.

Zur Validität der verwendeten Responseschwellen für die Responderanalysen zur Reduktion der Häufigkeit von Anfällen legt der pU keine Unterlagen vor. Es handelt sich um präspezifizierte Responseschwellen. Die Validität der Responseschwellen wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im zugrunde liegenden Anwendungsgebiet als valide eingeschätzt.

Die Validität hinsichtlich der Veränderung der anfallsfreien Tage erscheint weitestgehend adäquat. Es ist jedoch zu berücksichtigen, dass hierdurch keine reduzierte Krankheitslast durch eine verminderte Anfallshäufigkeit impliziert wird, da mehrere Anfälle an einem Tag kumuliert auftreten können. Hingegen kann die Zählweise des längsten anfallsfreien Intervalls bei fehlenden Werten (Nichtunterbrechung des Intervalls bei fehlenden Werten; siehe Operationalisierung) zu einer Überschätzung des anfallsfreien Intervalls führen, weshalb die Veränderung der anfallsfreien Tage gegenüber dem anfallsfreien Intervall als geeigneter angesehen und deshalb berücksichtigt wird.

Es werden somit die folgenden präspezifizierten Auswertungen als relevant für die vorliegende Bewertung eingestuft:

- prozentuale Veränderung der Häufigkeit jeglicher Anfälle pro 28 Tage gegenüber Baseline;
- Responderanalysen ($\geq 25\%$, $\geq 50\%$, $\geq 75\%$) zur Anzahl der Patienten mit einer Reduktion der Häufigkeit jeglicher Anfälle pro 28 Tage;
- prozentuale Veränderung der anfallsfreien Tage jeglicher Anfälle.

Der primäre Endpunkt zur prozentualen Veränderung schwerer motorischer Anfälle pro 28 Tage wird ergänzend dargestellt.

Globaler Betreuerereindruck (CGI)

Der Endpunkt „Globaler Betreuerereindruck“ mittels CGI-I wird in der Nutzenbewertung berücksichtigt. Die Endpunkte „Globaler Betreuerereindruck“ mittels CGI-CSID, CGICA, CGI-C–Zielverhalten werden aufgrund unklarer Validität und unklarer Patientenrelevanz nicht berücksichtigt.

Operationalisierung

Beschreibung

Der pU setzt die 4 folgenden CGI-Instrumente mit folgenden Fragestellungen zur Operationalisierung des Endpunktes ein, welche teilweise mehrere (inhaltlich identische) Versionen beinhalten:

- Caregiver Global Impression of Change in Seizure Intensity/Duration (CGI-CSID):
 - „Bewerten Sie die Veränderung der Anfallsintensität und/oder -dauer nach Beginn der Anwendung des Prüfpräparats.“
- Clinical Global Impression of Improvement (CGI-I; Caregiver and Clinician-Version):
 - „Bewerten Sie den globalen Gesamteindruck der Veränderung nach Beginn der Anwendung des Prüfpräparats im Vergleich zum Ausgangswert (vor Beginn der Anwendung des Prüfpräparats).“
- Caregiver Global Impression of Change in Attention (CGICA):
 - „Bewerten Sie die Veränderung der Aufmerksamkeit nach Beginn der Anwendung des Prüfpräparats.“
- Caregiver Global Impression of Change (CGI-C)–Zielverhalten:
 - „Bewerten Sie die Veränderung im Zielverhalten nach Beginn der Anwendung des Prüfpräparats.“

Alle vom pU eingesetzten CGI-Instrumente werden durch die Betreuungspersonen (einschließlich Eltern) bewertet.

Bei den zuletzt genannten Instrumenten (CGICA, CGI-C–Zielverhalten) erfolgt zur ersten Studienvsiste zudem die Definition bzw. die Festlegung auf ein konkretes auffälliges Verhalten, anhand dessen die Veränderung im Zeitverlauf gemessen werden soll. Beim CGI-C–Zielverhalten war zusätzlich einer der vordefinierten Bereiche „Geselligkeit“, „Kommunikation“, „Reizbarkeit“ oder „Hyperaktivität“ zu wählen, der das Verhalten der Patientin bzw. des Patienten am

stärksten beeinflusst. Laut Modul 4 und dem Studienprotokoll soll die Festlegung eines vorab zu definierenden Verhaltens auch beim CGI-I durchgeführt werden. Allerdings wird das festgelegte Verhalten als Zielverhalten beschrieben, so dass davon ausgegangen wird, dass es sich hierbei um den CGI-C handelt. Dort ist ein entsprechendes Feld für die Eintragung als Erinnerungshilfe vorgemerkt.

Die Beantwortung aller CGI-Instrumente erfolgte auf einer 7-Punkte-Skala, mit den Antwortoptionen: (1) „sehr stark verbessert“, (2) „stark verbessert“, (3) „leicht verbessert“, (4) „keine Veränderung“, (5) „leicht verschlechtert“, (6) „stark verschlechtert“, (7) „sehr stark verschlechtert“. Die Veränderung wurde beim CGI-I im Vergleich zum Zustand vor Beginn der Einnahme der Studienmedikation bewertet. Bei den anderen CGI-Versionen fehlt die Angabe des Referenzzeitraums innerhalb der Fragebögen.

Bewertung

CGI-I

Die Operationalisierung des Endpunktes CGI-I ist weitestgehend nachvollziehbar. Aus der Fragestellung zum „...globalen Gesamteindruck der Veränderung...“ ist ein expliziter Bezug zur Erkrankung, den davon hervorgerufenen Symptomen oder dem Gesundheitszustand nicht unmittelbar erkennbar. Es wird jedoch davon ausgegangen, dass der Bezug zur Studie und die damit verbundene globale Einschätzung des Gesundheitszustandes für die Betreuenden und Klinikerinnen bzw. Kliniker ersichtlich ist.

Aufgrund der angenommenen Nähe der Betreuungspersonen (meist Eltern) zu den Patientinnen und Patienten wird deren Einschätzung globaler Gesundheitsveränderungen als zuverlässiger angesehen; daher wird die Betreuerversion gegenüber der Klinikerversion bevorzugt dargestellt.

Weiterhin ist unklar, inwiefern beim CGI-I tatsächlich vorab ein Verhalten (analog zum CGICA, CGI-C–Zielverhalten) festgelegt werden musste, oder es sich hierbei um einen Fehler in der Beschreibung handelt.

CGI-CSID

Die Operationalisierung des Endpunktes CGI-CSID ist nicht nachvollziehbar. Innerhalb des Items werden mehrere Dimensionen (Dauer, Intensität) des Anfallsgeschehens mit einer und/oder Verknüpfung innerhalb einer Frage zusammenfasst (siehe Validität). Somit bleibt für die Auswertung des Instruments unklar, ob sich die Dauer der Anfälle, die Intensität der Anfälle oder beides verändert hat. Zudem gibt es inhaltliche Überschneidungen mit dem CGI-I, da diese Veränderungen auch über die globale Einschätzung der Veränderung beurteilt werden. Darüber hinaus ist die Referenzperiode, gegenüber derer die Veränderung beurteilt werden soll, im Fragebogen nicht beschrieben.

CGICA und CGI-C–Zielverhalten

Die Operationalisierung der übrigen CGI-C-Instrumente (CGICA und CGI-C–Zielverhalten) ist weitestgehend nachvollziehbar. Eine Abgrenzung zwischen dem Verhalten (CGICA) und der Aufmerksamkeit (CGI-C–Zielverhalten) scheint im vorliegenden Fall nicht uneingeschränkt möglich zu sein. Es bestehen inhaltliche Überschneidungen zwischen dem CGICA, dem CGI-C–Zielverhalten untereinander und gegenüber dem CGI-I (siehe Validität).

Darüber hinaus bleibt sowohl für den CGICA als auch für den CGI-C–Zielverhalten die jeweilige Referenzperiode, gegenüber derer die Veränderung beurteilt werden soll, unklar.

Generell fehlen zu allen CGI-I-Instrumenten Angaben, in welcher Reihenfolge diese administriert wurden. Es ist somit unklar, inwiefern zuvor ausgefüllte Instrumente möglicherweise einen Einfluss auf danach ausgefüllte Instrumente hatten. Auch ist nicht berichtet, inwiefern aus Konsistenzgründen die Erfassung des Endpunkts immer durch dieselbe Person erfolgen sollte.

Patientenrelevanz

Die globale Veränderung des Gesundheitszustandes mittel CGI-I wird per se als patientenrelevant erachtet.

Auch die Veränderung der Anfallsintensität und -dauer mittels des CGI-CSID wird als patientenrelevant erachtet, wobei jener die Aspekte nicht einzeln erfasst, da die Dauer der Anfälle, die Intensität der Anfälle oder beides bewertet wird. Aufgrund der nicht nachvollziehbaren Operationalisierung wird die Patientenrelevanz des CGI-CSID als unklar eingeschätzt.

Die Patientenrelevanz der mittels CGI-C-Zielverhalten beziehungsweise CGICA erfassten Verhaltensbereiche und des Aufmerksamkeitsverhaltens sind unklar. Beim CGI-C-Zielverhalten ist aufgrund der vorliegenden Ausgestaltung der vier auswählbaren Verhaltensbereiche – Sozialverhalten, Kommunikation, Reizbarkeit und Hyperaktivität – nicht hinreichend belegt, inwieweit die ausgewählten Bereiche das jeweils für die Personen relevante Spektrum des Verhaltens vollständig und angemessen abbilden. Zudem bleibt unklar, ob die verschiedenen auswählbaren Bereiche für die Betroffenen gleichermaßen relevant sind. Auch beim CGICA scheint unklar, welches konkrete Konstrukt im Zusammenhang mit „Aufmerksamkeit“ erfasst werden soll (z. B. Reaktionsfähigkeit, Wachheit oder Interaktionsfähigkeit). Darüber hinaus ist nicht hinreichend belegt, dass sich dieses Konstrukt eindeutig von den mittels CGI-C-Zielverhalten erfassten Verhaltensbereichen beziehungsweise vom individuell festgelegten Zielverhalten abgrenzen lässt.

Validität

Grundsätzlich ist die Validität aller vom pU eingereichten CGI-Instrumente unklar, da weder aus den Unterlagen des pU noch aus einer orientierenden Recherche psychometrische Validierungsstudien identifiziert werden konnten. Zudem ist bei allen CGI-Instrumenten außer dem CGI-I der Bezugszeitraum unklar, anhand dessen die Veränderung erfasst werden soll. Dennoch besteht beim CGI-I die Möglichkeit eines Recall-Bias aufgrund des 17-wöchigen DB-Studienzeitraumes.

CGI-I

In vorangegangenen Nutzenbewertungsverfahren wurden die Veränderungen im globalen klinischen Gesamtzustand bzw. Gesundheitszustand mittels CGI-I bereits als geeignet bewertet. Vor dem Hintergrund von kognitiven Beeinträchtigungen, einer verzögerten Entwicklung und dem überwiegend jungen Alter der Personen erscheint eine Fremdeinschätzung durch die betreuende Person an dieser Stelle angemessen. Jegliche Veränderung im Sinne einer Verbesserung des Gesundheitszustandes wird als relevant erachtet, auch wenn die Einschätzung des Gesundheitszustandes rein subjektiv durch Fremdbeurteilung erfolgt. Aufgrund der erforderlichen intensiven Betreuung der betroffenen Person ist davon auszugehen, dass hier eine adäquate Einschätzung durch die Betreuungsperson möglich ist.

CGI-CSID, CGI-C-Zielverhalten, CGICA

Die Validität der Instrumente CGI-CSID, CGI-C-Zielverhalten und CGICA kann nicht beurteilt werden. Aufgrund von Unklarheiten in der Operationalisierung und unklarer Patientenrelevanz werden die Endpunkte CGI-CSID, CGI-C-Zielverhalten und CGICA nicht zur Nutzenbewertung herangezogen.

Children's Sleep Habits Questionnaire (CSHQ)

Der Endpunkt CSHQ wird in der Nutzenbewertung nicht berücksichtigt aufgrund nicht-gegebener Validität.

Operationalisierung

Beschreibung

Der Children's Sleep Habits Questionnaire (CSHQ) erfasst das Schlafverhalten von Kindern und Jugendlichen anhand von 33 Items in 8 Domänen. Die Domänen umfassen:

- Probleme beim Zubettgehen (n = 6 Items; Items 1, 3–6, 8);
- Einschlafverzögerungen (n = 1 Item; Item 2);
- Schlafmenge (n = 3 Items; Items 9–11);
- Angstzustände im Zusammenhang mit dem Schlafen (n = 4 Items; Items 5, 7, 8, 21);
- Nächtliches Erwachen (n = 3 Items; Items 16, 24, 25);
- Parasomnien (n = 7 Items; Items 12–15, 17, 22, 23);
- Atmungsstörungen (n = 3 Items; Items 18–20) sowie
- Tagesmüdigkeit (n = 8 Items; Items 26–33).

Der Fragebogen wird fremdberichtet durch die Eltern oder Betreuungspersonen ausgefüllt. Der Bezugszeitraum beträgt 7 Tage.

Die Items 1 bis 31 werden auf einer dreistufigen Häufigkeitsskala erhoben: 1 = selten (0–1-mal pro Woche), 2 = manchmal (2–4-mal pro Woche), 3 = üblicherweise (5–7-mal pro Woche). Die Items 32 (Fernsehen) und 33 (Autofahren) erfassen den Grad der Schläfrigkeit des Kindes während dieser Aktivitäten: 1 = nicht schläfrig, 2 = sehr schläfrig, 3 = schläft ein. Zusätzlich enthält der Fragebogen zwei quantitative Fragen zur durchschnittlichen täglichen Schlafmenge in Stunden und zur Dauer nächtlicher Wachphasen in Minuten; diese gehen nicht in den Score ein.

Für die Auswertung werden die Werte der Items 1, 2, 3, 10, 11 und 26 umgekehrt, da bei diesen Items ein höherer Rohwert einem günstigeren Schlafverhalten entspricht. Der Score jeder Domäne ergibt sich als Summe der zugehörigen Items. Die Items 5 und 8 sind dabei zwei Domänen gleichzeitig zugeordnet: Sie gehen sowohl in die Domäne „Probleme beim Zubettgehen“ als auch in die Domäne „Angstzustände im Zusammenhang mit dem Schlafen“ ein. Die resultierenden Punktspannweiten der Domänen betragen: „Probleme beim Zubettgehen“ 6–18, „Einschlafverzögerungen“ 1–3, „Schlafmenge“ 3–9, „Angstzustände im Zusammenhang mit dem Schlafen“ 4–12, „Nächtliches Erwachen“ 3–9, „Parasomnien“ 7–21, „Atmungsstörungen“ 3–9 sowie „Tagesmüdigkeit“ 8–24. Der Gesamtscore wird als Summe aller 33 Items berechnet. Die Spannweite des Gesamtscores beträgt 33–99. Eine höhere Punktzahl in den Domänen und dem Gesamtscore entspricht einem schlechteren Schlafverhalten.

Bewertung

Die Operationalisierung ist eingeschränkt nachvollziehbar. Zwischen den Studienunterlagen besteht eine Inkonsistenz in der Domänen-Auswertung: Die Items 5 und 8 sind gemäß Originalpublikation [19] gleichzeitig der Domäne „Probleme beim Zubettgehen“ und der Domäne „Angstzustände im Zusammenhang mit dem Schlafen“ zugeordnet, sodass sich 35 Item-Positionen ergeben, während der Gesamtscore auf 33 distinkten Items basiert. Für den Gesamtscore ist die Auswertung damit adäquat. Für die Domänen-Scores der genannten beiden Domänen ist jedoch nicht nachvollziehbar, ob Items 5 und 8 validierungskonform in beide Domänen eingehen. Falls nicht, könnte dies zu einer systematischen Unterschätzung des Domänen-Scores führen.

Patientenrelevanz

Eine Verbesserung des Schlafverhaltens bildet eine patientenrelevante Veränderung der Morbidität ab [1, 13, 18]. Da Patientinnen und Patienten mit CDD aufgrund schwerer kognitiver und kommunikativer Einschränkungen keine Selbstauskunft erteilen können, wird die Fremdbeurteilung durch Eltern oder Betreuungspersonen als sachgerecht angesehen. Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

Validität

Der Children's Sleep Habits Questionnaire (CSHQ) ist ein elternberichtetes Screening-Instrument zur Erfassung von Schlafproblemen bei Kindern im Alter von 4 bis 10 Jahren [19]. Er umfasst 33 distinkte Items in acht Domänen, wobei zwei Items in jeweils zwei Domänen eingehen.

Die Originalvalidierung [19] erfolgte an einer Community-Stichprobe ($n = 469$, Ø 7,6 Jahre) und einer klinischen Stichprobe ($n = 154$, Ø 6,8 Jahre). Beide Validierungsstichproben schlossen Kinder unter Antikonvulsiva sowie Kinder mit psychiatrischen oder neurologischen Erkrankungen explizit aus. Die interne Konsistenz des Gesamtscores war akzeptabel ($\alpha = 0,68\text{--}0,78$), auf Domänenebene jedoch erheblich heterogen. Die Test-Retest-Reliabilität lag im moderaten bis guten Bereich ($r = 0,62\text{--}0,79$). Die deutsche Übersetzung (CSHQ-DE) [23] zeigte an einer Normstichprobe ($n = 298$) und einer klinischen Stichprobe ($n = 45$) vergleichbare Gütekriterien (Gesamtscore $\alpha = 0,68$; Domänen $\alpha = 0,23\text{--}0,70$; Retest $r = 0,76$). Eine Überprüfung der Konstruktvalidität [17] ergab keine signifikanten Korrelationen zwischen CSHQ-Domänen und Polysomnographie-Parametern; die Domänen unterschätzten objektiv vorhandene Schlafprobleme systematisch.

Für Kinder mit CDD liegt keine populationsspezifische Validierung vor. Die Validierungspopulation unterscheidet sich zudem wesentlich von der vorliegenden CDD-Studienpopulation, da Kinder mit neurologischen Erkrankungen und Antikonvulsiva-Therapie – und somit die Population der vorliegenden Studie – explizit ausgeschlossen wurden. Die Übertragbarkeit der dort untersuchten psychometrischen Eigenschaften auf Kinder mit CDD ist daher nicht belegt. Zweitens liegt bei wesentlichen Items eine inhaltliche Fehlpassung (mangelnde Augenscheinvalidität) vor, denn mehrere Items können krankheits- oder medikationsbedingte Eigenschaften statt des intendierten Schlafkonstrukts erfassen und dadurch Bodeneffekte verursachen. Zuletzt wurde der CSHQ als einmaliges Querschnitts-Screening konzipiert. Eine Validierung der Änderungssensitivität für den Einsatz als Verlaufsmaß in klinischen Studien fehlt.

Der CSHQ wird daher für die vorliegende CDD-Population als nicht ausreichend validiert angesehen.

Anxiety, Depression and Mood Scale (ADAMS)

Der Endpunkt ADAMS wird in der Nutzenbewertung nicht berücksichtigt aufgrund nicht-gegebener Validität.

Operationalisierung

Beschreibung

Der Anxiety, Depression and Mood Scale (ADAMS) erfasst Angstzustände, Depressionen und Stimmungsschwankungen bei Menschen mit geistiger Behinderung anhand von 28 Items in fünf Domänen. Die Domänen umfassen:

- Manisches/hyperaktives Verhalten ($n = 5$ Items; Items 3, 4, 12, 17, 22);
- Depressive Stimmung ($n = 7$ Items; Items 5, 9, 10, 14, 18, 23, 28);
- Soziale Vermeidung ($n = 7$ Items; Items 2, 6, 13, 19, 21, 25, 27);
- Allgemeine Angst ($n = 7$ Items; Items 1, 3, 7, 11, 15, 24, 26) sowie
- Obsessives/zwanghaftes Verhalten ($n = 3$ Items; Items 8, 16, 20).

Der Fragebogen wird fremdberichtet durch die Eltern oder Betreuungspersonen ausgefüllt und basiert auf dem Verhalten der betroffenen Person.

Die Items werden auf einer vierstufigen Likert-Skala bewertet: 0 = das Verhalten ist nicht aufgetreten oder stellt kein Problem dar, 1 = das Verhalten tritt gelegentlich auf oder stellt ein leichtes Problem dar, 2 = das Verhalten tritt recht häufig auf oder stellt ein moderates Problem dar, 3 = das Verhalten tritt häufig auf oder entspricht einem schweren Problem.

Der Score jeder Domäne ergibt sich als Summe der zugehörigen Items. Item 3 ist dabei zwei Domänen gleichzeitig zugeordnet: Es geht sowohl in die Domäne „Manisches/hyperaktives Verhalten“ als auch in die Domäne „Allgemeine Angst“ ein. Die resultierenden Punktskizzen der Domänen betragen: „Manisches/hyperaktives Verhalten“ 0–15, „Depressive Stimmung“, „Soziale Vermeidung“ und „Allgemeine Angst“ 0–21 sowie „Obsessives/zwanghaftes Verhalten“ 0–9. Für die vorliegende Auswertung wird kein Gesamtscore berechnet. Eine höhere Punktzahl in den Domänen entspricht einem schlechteren Verhalten in Bezug auf Angstzustände, Depressionen und Stimmungsschwankungen.

Der Bezugszeitraum ist nicht einheitlich: Zur Baseline-Visite werden die Betreuungspersonen gebeten, das Verhalten der betroffenen Person über die letzten 6 Monate zu beurteilen. Bei allen Folgevisiten bezieht sich die Beurteilung auf den Zeitraum seit der letzten Visite; die Abstände zwischen den Folgevisiten in der DBP betragen 4 Wochen (Woche 5 bis Woche 9) bzw. 8 Wochen (Woche 9 bis Woche 17).

Bewertung

Die Operationalisierung ist nachvollziehbar. Allerdings wird der Bezugszeitraum von 6 Monaten zur Baseline-Visite methodisch problematisch eingeschätzt, da er sich grundlegend von den Bezugszeiträumen der Folgevisiten (4 bzw. 8 Wochen) unterscheidet. Während zu Baseline ein halbjährlicher Rückblick verlangt wird, beziehen sich die Folgeerhebungen auf einen deutlich kürzeren Zeitraum.

Patientenrelevanz

Verhaltensauffälligkeiten und psychiatrische Komorbiditäten zählen bei der CDD neben der therapierefraktären Epilepsie und der schweren Entwicklungsverzögerung zu den zentralen und klinisch bedeutsamen Krankheitsmanifestationen [2]. Da Patientinnen und Patienten mit CDD aufgrund schwerer kognitiver und kommunikativer Einschränkungen keine Selbstauskunft erteilen können, wird die Fremdbeurteilung durch Eltern oder Betreuungspersonen als methodisch sachgerecht angesehen. Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

Validität

Die Anxiety, Depression And Mood Scale (ADAMS) ist ein fremdbeurteilendes Screening-Instrument zur Erfassung von Angst-, Depressions- und Stimmungssymptomen bei Personen mit geistiger Behinderung [6]. Das Instrument umfasst 28 Items in 5 Domänen: Manisches/hyperaktives Verhalten, Depressive Stimmung, Soziale Vermeidung, Allgemeine Angst und Zwanghaftes Verhalten.

Die Originalvalidierung [5] erfolgte an zwei Stichproben mit geistiger Behinderung (Sample 1: n = 265; Sample 2: n = 268; Altersbereich 10–79 Jahre). Die interne Konsistenz der Domänen war gut (Cronbachs Alpha = 0,75–0,83), die Konstruktvalidität (konfirmatorische Faktorenanalyse) fiel mit einem RMSEA (Root Mean Square Error of Approximation) von 0,087 und NNFI (Non-Normed Fit Index) von 0,75 jedoch lediglich mäßig aus. Die Interrater-Reliabilität für die Domänen lag bei 0,37–0,62. Weitere Validierungsstudien [8, 21] bestätigten eine gute interne Konsistenz und Test-Retest-Reliabilität an erwachsenen Stichproben.

Für das Anwendungsgebiet der CDD liegt keine populationsspezifische Validierungsstudie vor. Die Validierungspopulation unterscheidet sich zudem wesentlich von der vorliegenden CDD-Studienpopulation, da diese einen Altersbereich von 10 bis 79 Jahren (Durchschnitt 38,7 Jahre) umfasst, während die Marigold-Studie bereits Kinder ab 2 Jahren einschließt; für Kinder unter 10 Jahren liegen damit keinerlei Validierungsdaten vor. Zweitens besteht eine inhaltliche Fehlpassung (mangelnde Augenscheinvalidität) des Instruments. Da CDD-Patienten in der Regel keine verbale Sprache entwickeln und ein durch die Epilepsie geprägtes Verhaltensprofil aufweisen, ist die Fremdbeurteilung innerer Zustände wie Angst oder Traurigkeit nicht valide möglich. Symptome wie häufiges Weinen oder fehlende Mimik spiegeln bei CDD häufig neurologische Charakteristika oder postiktale Zustände wider und keine psychiatrischen Symptome. Zudem wurde die ADAMS als Querschnitts-Screening-Instrument entwickelt; eine Validierung der Änderungssensitivität sowie ein MCID (Minimal Clinically Important Difference) fehlen. Der geforderte sechsmonatige Bezugszeitraum ist zudem anfällig für Erinnerungsverzerrungen (Recall Bias), was die Messgenauigkeit als Verlaufsmaß in klinischen Studien zusätzlich gefährdet.

Die ADAMS wird für den Einsatz in der vorliegenden Population daher als nicht ausreichend validiert angesehen. Eine belastbare Interpretation der erhobenen Ergebnisse ist nicht möglich.

2.3.3 Lebensqualität

Quality of Life Inventory-Disability (QI-Disability)

Der Endpunkt Quality of Life Inventory-Disability (QI-Disability) wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

In der Studie Marigold wird der QI-Disability-Fragebogen zur Erhebung der gesundheitsbezogenen Lebensqualität bei Kindern und jungen Erwachsenen ab 2 Jahren eingesetzt. Der QI-Disability ist ein für Kinder mit geistigen Behinderungen entwickeltes Instrument, das als Fremdbewertung durch Betreuungspersonen bzw. Eltern konzipiert ist. Es umfasst 32 Items, welche sich auf die folgenden 6 Domänen verteilen:

- Physische Gesundheit (Items 1–4)
- Positive Emotionen (Items 5–8)
- Negative Emotionen (Items 9–15)
- Soziale Interaktion (Items 16–22)
- Freizeit und Outdoor-Aktivitäten (Items 23–27)
- Unabhängigkeit (Items 28–32)

Jedes Item wird auf einer 5-Punkte-Skala beantwortet: (1) „nie“, (2) „selten“, (3) „manchmal“, (4) „oft“, (5) „sehr oft“. Zur Scorebildung, werden zuerst die Itemwerte in Scores überführt, die einen Wert von 0 bis 100 einnehmen, wobei höhere Werte eine bessere Lebensqualität anzeigen. Die 6 Domänenscores werden durch Invertierung der negativen Emotionen ($6 - \text{Rohwert}$) und anschließender Mittelung der Items innerhalb der jeweiligen Domäne gebildet ($25 \times (\text{Rohwert} - 1)$); der Gesamtscore errechnet sich durch anschließende Mittelung der jeweiligen Domänenscores. Der Bezugszeitraum des QI-Disability umfasst den letzten Monat.

Bewertung

Die Operationalisierung ist nachvollziehbar.

Es fehlen Angaben, in welcher Reihenfolge die Instrumente administriert wurden. Es ist daher unklar, ob und inwiefern die möglicherweise zuvor erhobenen Instrumente Einfluss auf das Ausfüllen des QI-Disability hatten.

Patientenrelevanz

Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen.

Validität

Der QI-Disability ist ein fremdbeurteilender Fragebogen für Kinder und Jugendliche mit geistigen und körperlichen Einschränkungen (u. a. Down-Syndrom, Rett-Syndrom, Zerebralparese, Autismus-Spektrum-Störung), der in mehreren Studien entwickelt und validiert wurde [3, 11]. Das Instrument wurde zudem bereits in einer Studie mit CDD-Patientinnen und Patienten (3–11 Jahre: 61 %, 12+ Jahre: 39 %; weiblich: 82 %) eingesetzt [12] und innerhalb dieser psychometrisch überprüft [22]. Die Items und Domänenstruktur wurden iterativ auf Basis qualitativer Studien mit pflegenden Eltern (semistrukturierte Interviews) 5–18-jähriger Kinder entwickelt; der initiale 50-Item-Fragebogen wurde im Anschluss schrittweise durch kognitive Interviews (Alter: 6–18-Jahre) und Faktorenanalysen auf 32 Items reduziert [4]. Die Items des Fragebogens decken die Dimensionen der Lebensqualität (physisch, psychisch, sozial) ab, sodass insgesamt die Inhaltsvalidität als gegeben angesehen wird.

Innerhalb der Originalstudie [3] erfolgte die Überprüfung der psychometrischen Eigenschaften anhand von 253 intellektuell eingeschränkten Kindern (Down-Syndrom, Rett-Syndrom, Zerebralparese und Autismus-Spektrum-Störung) im Alter von 5 bis 18 Jahren ($\bar{x}$: 12,2 Jahre; weiblich: 62,1 %). Eine der weiteren Validierungsstudien umfasste ein ähnliches Patientenkollektiv (n: 60, $\bar{x}$: 11,5 Jahre, Spanne: 5–19 Jahre, weiblich: 56,4 %) bei gleichen Krankheitsbildern [11], die zweite erfolgte an CDD-Patientinnen und Patienten (n: 152 bzw. 56; 3–11 Jahre: 61%, 12+ Jahre: 39%; weiblich: 82%) [22].

Die strukturelle Validität in der initialen Studie sowie in der Studie der CDD-Population ist unter Berücksichtigung der Cut-off-Werte insgesamt akzeptabel, in einzelnen Analysen werden die gewünschten Grenzwerte jedoch leicht unterschritten bzw. überschritten [3, 22].

Die konvergente Validität ist eingeschränkt: Nur 2 von 6 Domänen erreichen den geforderten Average-variance-extracted-(AVE-)Cut-off von $>0,5$ („Positive Emotionen“, „Freizeit und Outdoor-Aktivitäten“); die übrigen Domänen liegen bei $AVE \leq 0,5$, was auf eine begrenzte Trennschärfe der Konstrukte hindeutet [3]. Dies spiegelt sich hingegen nicht in der CDD-Population wider (alle Domänen $AVE > 0,5$) [22]. Die divergente Validität ist überwiegend akzeptabel [3], zeigt jedoch insbesondere in den Domänen „Soziale Interaktion“ (0,67 vs. 0,71) und „Unabhängigkeit“ (0,58 vs. 0,71) in der CDD-Population Einschränkungen [22]. Die Known-Groups-Validität wird durch erwartungskonforme Unterschiede zwischen Diagnosen, Funktionsniveaus und Altersgruppen gestützt (z. B. höhere Unabhängigkeitswerte bei Down-Syndrom vs. Rett-Syndrom, niedrigere soziale Interaktion bei Autism Spectrum Disorder (ASD)) [3].

Die interne Konsistenz (Reliabilität) ist in allen Domänen studienübergreifend moderat bis hoch und ergibt Cronbach's Alpha-Werte im Bereich von 0,72 bis 0,90, bzw. 0,68 bis 0,90 [11] und 0,71 bis 0,91 [22], wobei die Domäne „Physische Gesundheit“ jeweils die niedrigsten und „Positive Emotionen“ die höchsten Werte zeigt. Die Test-Retest-Reliabilität ist überwiegend gut (domänenspezifischer Intraclass Correlation Coefficient (ICC) meist $>0,7$), mit durchgehend höheren Werten für den Gesamtscore ($>0,9$) als für die Domänenscores; schwächere Übereinstimmungen zeigen sich für die Domänen „Physische Gesundheit“ (0,51 roh; 0,68 adjustiert) und „Positive Emotionen“ (0,50 bzw. 0,58) [11, 22]. Es gibt erste – über 2 Validierungsstudien hinweg konsistente – Hinweise auf eine vorhandene Responsiveness für den Gesamtscore und einzelne Domänen (u. a. physische Gesundheit) gegenüber Veränderungen der kindlichen Gesundheit und des Verhaltens. In weiteren Domänen unterschieden sich die Ergebnisse zwischen den Studien jedoch [11, 22]. Die Responsiveness wurde ankerbasiert untersucht. Boden- oder Deckeneffekte werden nicht berichtet.

Der QI-Disability wurde ursprünglich an Kindern und Jugendlichen von 5 bis 19 Jahren entwickelt und validiert [3, 11]. In der CDD-Population fand zudem eine Validierung an 3–12- und ≥ 12 -jährigen Patientinnen und Patienten (61,2 % bzw. 38,8 %) statt, wobei das Alter und die Altersobergrenze der ≥ 12 -Jährigen aufgrund fehlender Angaben zur Altersverteilung nicht weiter spezifiziert werden. Da die Marigold-Studie 2–21-jährige Personen einschließt, liegt für 2-jährige Patientinnen und Patienten keine Evidenz und für die 3–4-Jährigen nur spezifische Evidenz zur Messgüte im Kontext von CDD vor. Für Kinder und junge Erwachsene (≥ 12 Jahre) liegt aufgrund der fehlenden Informationen zudem eine unklare Evidenz spezifisch für die CDD-Population vor. Insbesondere für die jüngsten Patientinnen und Patienten können daher einzelne Fragen (z. B. Frage zum Gebrauch technischer Geräte oder zu Freizeitaktivitäten) weniger geeignet sein.

Zusammenfassend wird der QI-Disability trotz gewisser Limitationen (konvergente/divergente Validität, Test-Retest-Reliabilität in bestimmten Domänen) als weitgehend valide und reliabel angesehen. Auch in Hinblick auf die Population stimmen die Entwicklungs- und die anschließenden Validierungsstudien mit der vorliegenden Indikation überwiegend überein (mit Ausnahme der 2- bzw. 2–4-jährigen Kinder), weshalb der QI-Disability zur Nutzenbewertung herangezogen wird. Die Einschätzung mittels Fremdbewertung erscheint aufgrund der kognitiven Einschränkungen und des geringen Alters der Betroffenen adäquat.

2.3.4 Sicherheit

Unerwünschte Ereignisse

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

Operationalisierung

Beschreibung

Ein unerwünschtes Ereignis (UE) umfasst in der Studie Marigold jegliche ungünstigen und unbeabsichtigten Anzeichen (einschließlich anormaler Laborbefunde), Symptome oder Erkrankungen, unabhängig davon, ob ein kausaler Zusammenhang mit dem Studienmedikament besteht oder nicht. Erhoben werden UE ab dem Tag der Einwilligung bis 2 Wochen nach Abschluss der Taper-Phase und beinhalten somit auch die 6-wöchige Baseline-Phase.

Berücksichtigt werden für die Sicherheitspopulation alle UE, die im verblindeten Zeitraum nach der ersten Medikationsgabe bis zu 2 Wochen nach Abschluss der Taper-Phase (aufgrund des Studienabschlusses oder eines vorzeitigen Abbruchs) entweder neu auftraten oder sich verschlechterten (Treatment Emergent Adverse Event (TEAE)). Wurde ein UE für dieselbe Person mehr als einmal gemeldet, wird diese Person bei der Inzidenz für dieses UE nur einmal gezählt. Bei Personen, bei denen dasselbe UE in mehreren Schweregraden auftrat, wird bei der Analyse der Inzidenz nach Schweregrad das Auftreten des UE mit dem höchsten Schweregrad herangezogen.

Die standardisierte Kodierung von UE erfolgte mittels „Medical Dictionary for Regulatory Activities“ (MedDRA; Version 23.0) nach Systemorganklasse (SOC) und Preferred Term (PT).

Die Klassifikation der UE nach Schweregrad erfolgte anhand folgender studieneigener Kriterien:

- Leicht: UE, die leichte oder vorübergehende Beschwerden auslösen und eine minimale therapeutische Intervention oder Behandlung erfordern. Die alltäglichen Aktivitäten werden nicht beeinträchtigt.
- Moderat: UE, die tägliche Aktivitäten einschränken oder beeinträchtigen und durch spezifische therapeutische Interventionen gelindert werden können. Sie stellen jedoch kein signifikantes oder permanentes Risiko für die Patientinnen und Patienten dar.
- Schwer: UE, die zu erheblichen Einschränkungen des Gesundheitszustandes und Aktivitäten des täglichen Lebens führen und eine intensive therapeutische Intervention erfordern.

Als schwerwiegendes unerwünschtes Ereignis (SUE) wurden Ereignisse definiert, die zu einem der folgenden Ereignisse führten:

- Tod;
- Lebensbedrohung;
- stationäre Krankenhausaufnahme oder Verlängerung eines bestehenden Krankenhausaufenthalts;
- anhaltende oder erhebliche Behinderung bzw. Unfähigkeit;
- angeborene Fehlbildung oder Geburtsfehler;
- Risiko eines schwerwiegenden Ereignisses, das eine medizinische/chirurgische Präventionsmaßnahme erforderte.

Im Rahmen einer Sensitivitätsanalyse wurden Post-hoc-Auswertungen unter Nichtberücksichtigung von epileptischen Anfällen als krankheitsbezogenes Ereignis vorgelegt.

Bewertung

Die Operationalisierung ist weitestgehend nachvollziehbar. Eine Begründung für die studien-spezifische Einteilung der Schweregrade führt der pU nicht an, jedoch ist die Einteilung an die Standardklassifikation "Common Terminology Criteria for Adverse Events" (CTCAE) angelehnt und wird als adäquat eingestuft. Der pU reicht post hoc eine Analyse unter Nichtberücksichtigung von epileptischen Anfällen als Progressionsereignis ein. Eine Definition für die spezifischen Anfallsereignisse wird nicht vorgelegt, zudem wurden keine weiteren krankheitsbezogenen Ereignisse herausgerechnet. Die Analyse wird für die vorliegende Nutzenbewertung herangezogen. Es ist jedoch nicht auszuschließen, dass weitere Symptome der Grunderkrankung in die UE-Erfassung eingehen.

Patientenrelevanz

Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen. Die Patientenrelevanz von Laborparametern ist unklar.

Validität

Die Erhebung der unerwünschten Ereignisse wird in der vorliegenden Studie grundsätzlich als valide eingeschätzt.

2.3.5 Erhebungszeitpunkte

Eine Übersicht der Erhebungszeitpunkte der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte findet sich in Tabelle 6.

Tabelle 6: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie Marigold

Endpunkt \ Studienvisite Visite ¹⁾ Woche	Screening		Baseline	DBP (Titrations- und Erhaltungsperiode)		Finale DB-Visite	Taper- Visite ²⁾
	- ³⁾ -14	1 -6	2 ⁴⁾ 0	3 5	4 9	5 17	- -
Todesfälle ⁵⁾	kontinuierlich						
Epileptische Anfälle ⁶⁾	kontinuierlich ⁷⁾⁸⁾						
CGI-I		⁹⁾		X	X	X	X
QI-Disability			X	X	X	X	X
Unerwünschte Ereignisse	kontinuierlich						

- ¹⁾ Zusätzlich zu den Visiten vor Ort erfolgten telefonische Anrufe zur Prüfung der Nebenwirkungen der Studienmedikation und zur Erfassung der Compliance der Anfallserfassung mittels elektronischem Anfallstagebuch an Tag 3 nach Baseline sowie wöchentlich an Woche 1 bis 4 sowie an Woche 11 bis 13 der DBP.
- ²⁾ Personen, die die Studie entweder vorzeitig abgebrochen oder beendet haben und nicht an der Open-Label-Extensionsstudie teilnehmen, durchliefen eine 2-wöchige Taper-Periode.
- ³⁾ Historische Anfallsdatenerhebung (Woche -14 bis Woche -6). Hier erfolgte das Einholen der Einwilligungserklärung zur Studienteilnahme.
- ⁴⁾ Visite 2 entspricht dem Baseline- und Randomisierungszeitpunkt; Visite 2 erfolgte mindestens 6 Wochen (42 Tage) nach Visite 1 und maximal nach 6 Wochen + 6 Tagen (48 Tage).
- ⁵⁾ Todesfälle wurden im Rahmen der Sicherheitserhebung erfasst.
- ⁶⁾ In verschiedenen Operationalisierungen vorliegend. Der primäre Endpunkt (prozentuale Veränderungen der Häufigkeit der schweren motorischen Anfälle) wird ergänzend dargestellt.
- ⁷⁾ Die Erhebung der 17-wöchigen DBP endet am Tag vor Visite 5 bzw. am Tag mit den letzten verfügbaren Anfallsdaten; bei Abschluss ohne Visite 5 und ohne Eintritt in die OLE endet die Erhebung am Tag vor der Ausschleichvisite.
- ⁸⁾ Die Erhebung mittels elektronischem Anfallstagebuchs erfolgte täglich durch die Betreuenden.
- ⁹⁾ Laut Studienprotokoll und Modul 4 ist die Erhebung sowohl für den CGICA, den CGI-C als auch für den CGI-I vorgesehen. Dies erscheint nicht korrekt, da der CGI-I, im Gegensatz zum CGICA bzw. CGI-C, die globale allgemeine Verbesserung (nicht spezifisch das Verhalten) abbildet. Es wird davon ausgegangen, dass keine Erhebung des CGI-I zu Visite 1 stattfand.

Abkürzungen: CGI-C: Caregiver Global Impression of Change; CGI-I: Clinical Global Impression of Improvement; CGICA: Caregiver Global Impression of Change in Attention; DBP: doppelblinde, placebokontrollierte Studienphase; DB-Visite: Datenbank-Visite; OLE: offene einarmige Studienphase (Open-Label Extension); QI-Disability: Quality of Life Inventory-Disability.

2.4 Statistische Methoden

Die statistischen Auswertungen basieren auf dem SAP 5.0 der Studie Marigold vom 18.05.2021. Alle SAP-Versionen wurden erst nach Studienstart (30.06.2018) finalisiert (SAP-Version 1.0 vom 22.08.2019). Die SAP-Version 4.0 (30.03.2021) lag zeitlich zudem zwischen dem Datenschnitt zum Datenbankschluss (01.09.2020) und dem Datenschnitt nach Datenbankupdate (19.05.2021).

Analysepopulationen

- Intention-to-Treat (ITT): Beinhaltet alle Personen, die auf die Studienarme randomisiert wurden und mindestens eine Dosis der Studienmedikation erhalten haben. Die ITT-Population stellt die primäre Analysepopulation für alle Wirksamkeitsanalysen dar.
- Sicherheitspopulation: Umfasst alle auf die Studienarme randomisierten Personen, die mindestens eine Dosis der Studienbehandlung erhalten haben. Patientendaten wurden nach erhaltener Behandlung ausgewertet. Die Sicherheitspopulation entspricht der ITT-Population.

Datenschnitte

Es liegen Datenschnitte innerhalb der 17-wöchigen DBP vor:

- Datenschnitte zum Ende der DBP vom 31.07.2020 bzw. 01.09.2020
- Datenschnitt zum Ende der DBP (präspezifiziert) nach einem Datenupdate vom 19.05.2021

Alle im vorliegenden Dossier und in der CSR-Version 2.0 berichteten Auswertungen zur doppelblinden Behandlungsphase beziehen sich auf den Datenstand vom 19.05.2021. Für diese nachträglichen Datenaktualisierungen liegen weder eine Rationale noch weitere Spezifikationen zu den vorgenommenen Anpassungen vor. Innerhalb der CSR-Versionen 1.0 (12.01.2021) und 2.0 (28.05.2021) zeigen sich für den Endpunkt „epileptische Anfälle“ (einschließlich des primären Endpunktes "schwere motorische Anfälle") geringfügige Abwei-

chungen in den Ergebnissen, die tendenziell zu kleineren Unterschieden zwischen den Behandlungsarmen führen. Statistisch signifikante Änderungen liegen nicht vor, sodass von keinem verzerrenden Einfluss ausgegangen wird. Für die Nutzenbewertung wird der präspezifizierte Datenschnitt zum Ende der DBP mit den aktualisierten Daten vom 19.05.2021 herangezogen.

Präspezifizierte Subgruppenanalysen

Es liegen keine für die Nutzenbewertung relevanten Subgruppenauswertungen vor.

Präspezifizierung geplanter und durchgeführter Analysen

Epileptische Anfälle

Für die Erhebung der Anfälle musste im Tagebuch täglich vermerkt werden, ob ein Anfall aufgetreten ist oder nicht. Für die Berechnung der prozentualen Veränderung der medianen 28-tägigen Anfallshäufigkeit in der 17-wöchigen DBP gegenüber der 6-wöchigen Baseline-Phase wurde die folgende Formel verwendet:

$$\left(\frac{(\text{Anfallshäufigkeit pro 28 Tage}) - (\text{Anfallshäufigkeit pro 28 Tage zu Baseline})}{(\text{Anfallshäufigkeit pro 28 Tage zu Baseline})} \right) * 100 \%$$

Die Auswertung zur Häufigkeit epileptischer Anfälle erfolgte prädefiniert deskriptiv unter Angabe der arithmetischen Mittelwertsveränderung gegenüber Baseline. Für den primären Endpunkt (prozentuale Veränderung der Anfallshäufigkeit schwerer motorischer Anfälle) wurden zur Beurteilung des Gruppenunterschiedes (mediane Differenz zwischen den Behandlungsarmen) zudem nichtparametrische Verfahren angewendet (Hodges-Lehmann-Schätzer als Effektmaß und Wilcoxon-Rangsummentest zur Bestimmung des p-Wertes sowie verteilungsfreie 95%-Konfidenzintervalle (KI)). Post hoc wurden diese Analysen ebenfalls für jegliche Anfälle vorgelegt.

Die Analyse der Responder erfolgte für die prädefinierten Responderraten ($\geq 25 \%$, $\geq 50 \%$, $\geq 75 \%$) für die schweren motorischen und jeglichen Anfälle ebenfalls deskriptiv.

Der Gruppenvergleich bezogen auf die $\geq 50\%$ -Responderschwelle erfolgt zudem mittels Fisher's exaktem Test (95%-KI mittels Wald-Methode) sowie durch Darstellung einer kumulativen Responderkurve. In Modul 4 wurde als Effektmaß das Relative Risiko (RR) auch für die jeglichen Anfälle und das Signifikanzniveau (mittels Fisher's exaktem Test) berichtet. Fehlende Werte gehen nicht in die Analyse ein.

Die Berechnung der anfallsfreien Tage in der 17-wöchigen DBP (Anzahl der anfallsfreien Tage $\div$ Gesamtanzahl der Tage mit Anfallsdaten $\times 100$) erfolgte ebenfalls über die mediane Differenz (Hodges-Lehmann-Schätzer); post hoc wurde der p-Wert (Wilcoxon-Rangsummentest) ergänzt. Ausschließlich post hoc erfolgte zudem die Berechnung der anfallsfreien Tage ohne schwere motorische Anfälle.

Aufgrund der post hoc berichteten Effektmaße und des Signifikanzniveaus, gegenüber den prädefiniert überwiegend deskriptiven Analysen, werden die in Modul 4 des Dossiers präsentierten Analysen für die jeglichen Anfälle (prozentuale Veränderung der Anfallshäufigkeit, anfallsfreie Tage, Responderanalysen) berücksichtigt. Als Effektmaß für die Responderanalysen wird das Relative Risiko (RR) berichtet.

CGI

Die CGI-Instrumente wurden prädefiniert deskriptiv ausgewertet. Für den CGI-I erfolgte zusätzlich eine ordinale logistische Regression (mittels Proportional-Odds-Modell) mit der Behandlungsgruppe als erklärenden Faktor, unter Angabe von OR, 95%-KI und p-Wert.

In Modul 4 wurden post hoc zudem Responderanalysen zur jeglichen Verbesserung bzw. zur Verschlechterung (unter Angabe des RR) sowie kontinuierliche Analysen mittels MMRM eingereicht.

QI-Disability

Die prädefinierte Auswertung des QI-Disability erfolgte deskriptiv über Mittelwertdifferenzen zwischen den Gruppen zu den jeweiligen Erhebungszeitpunkten. Mit Modul 4 wurden post hoc Responderanalysen (unter RR-Angabe, 95%-KI und p-Wert) zur Verbesserung um $\geq 15\%$ der Skalenspannweite eingereicht. Responderanalysen zur Verschlechterung wurden nicht vorgelegt. Fehlende Werte wurden nicht berücksichtigt und nicht als Non-Responder gewertet. Zudem wurden kontinuierliche Analysen mittels MMRM vorgelegt. Analog zum CGI-I werden die post hoc berichteten Responderanalysen dargestellt.

Unerwünschte Ereignisse

Die deskriptive Auswertung der Inzidenzen war präspezifiziert. Post hoc wurde für Modul 4 als Effektmaß das RR sowie das Signifikanzniveau (p-Wert) berichtet.

Fehlende Werte und Daten-Imputation

Fehlende Werte wurden in der DBP grundsätzlich nicht imputiert. Fehlende Werte wurden als Missing Completely at Random angenommen. Fehlte ein Item eines Fragebogens, wurden auch die daraus berechneten Domänenscores sowie der Gesamtscore als fehlend behandelt. Ausnahme sind die nachfolgend dargestellten Sensitivitätsanalysen, im Rahmen deren eine Datenimputation erfolgte. Die Sensitivitätsanalysen waren für den primären Endpunkt präspezifiziert und wurden post hoc in gleicher Methodik für sämtliche Endpunkte durchgeführt. Insgesamt wurden 3 Sensitivitätsanalysen präspezifiziert, post hoc werden hingegen nur Sensitivitätsanalyse 1 und 2 berichtet:

1. Sensitivitätsanalyse

Fehlende tägliche Anfallsdaten während der DBP wurden durch den Median entsprechender Tage im Placeboarm ersetzt. Die Imputation erfolgte stratifiziert nach 4 Baseline-Quartilen der Anfallshäufigkeit, um die unterschiedliche Krankheitsintensität der Patientinnen und Patienten zu berücksichtigen und fehlende Werte möglichst durch Daten vergleichbarer Personen zu ersetzen.

2. Sensitivitätsanalyse

Fehlende tägliche Anfallsdaten wurden durch den Median der 5 höchsten Anfallszahlen entsprechender Tage im Placeboarm ersetzt, um die Annahme einer höheren Anfallshäufigkeit bei vorzeitigem Abbruch der Tagebucheinfassung zu prüfen.

Bei vorzeitigem Studienabbruch werden die zur Taper-Visite durchgeführten Erhebungen für die Analyse der Visite zugeordnet, deren geplanter Studientag innerhalb von 32 Tagen des tatsächlichen Tages der Taper-Visite liegt. Liegt der Tag außerhalb dieses 32-Tage-Fensters aller weiteren geplanten Visiten, erfolgt keine Zuordnung zu einer regulären Visite. Nicht zugeordnete Daten der Taper-Visite fließen nicht in die Auswertungen ein, werden aber in den Datenlistings aufgeführt. Die Datenlistings liegen nicht vor.

Einschätzung der statistischen Auswertungen

Die angewendeten statistischen Methoden erscheinen grundsätzlich angemessen. Für die Nutzenbewertung werden die post hoc eingereichten Responderanalysen für den CGI-I zur jeglichen Verbesserung (Summe aus „geringfügig verbessert“, „stark verbessert“ und „sehr stark verbessert“) und QI-Disability (Verbesserung um mindestens 15 % der Skalenspannweite) herangezogen. Dies wird aufgrund des eingeschränkten Verschlechterungspotenzials in der zugrundeliegenden eher schwer betroffenen CDD-Studienpopulation und des nicht eindeutig progredienten Krankheitsverlaufs als sachgerecht angesehen und dargestellt. Zu berücksichtigen ist, dass Personen mit fehlenden Werten aus der jeweiligen Auswertung ausgeschlossen und nicht als Non-Responder gewertet wurden.

Der teilweise hohe Anteil fehlender Werte in den Behandlungsarmen (CGI-I = 14 % bzw. 8 %; QI-Disability = jeweils 16 % im Gesamtscore) ist zu berücksichtigen. Der Anteil fehlender Werte ist nur im Placebo-Arm bei der CGI-I-Erhebung vollständig über Studienabrechende (4 % bzw. 8 %) erklärbar, ansonsten sind die Gründe für das Fehlen unklar. Diese Personen mit fehlenden Werten wurden zudem aus der jeweiligen Auswertung ausgeschlossen und nicht als Non-Responder gewertet, was grundsätzlich kritisch zu bewerten ist. Bei den post hoc eingereichten QI-Disability-Responderanalysen wurde zudem nur eine Verbesserung der Skalenspannweite, nicht jedoch eine entsprechende Verschlechterung vorgelegt. Analog zum CGI-I wird diese Fokussierung aufgrund des größeren klinischen Potenzials für Verbesserungen jedoch als sachgerecht angesehen.

Anzumerken ist, dass weder eine stratifizierte Randomisierung noch adjustierte Analysen unter Berücksichtigung der Anfallshäufigkeit zu Baseline durchgeführt wurden. Die Auswertungen (u. a. zur prozentualen Veränderung der Anfallshäufigkeiten epileptischer Anfälle) basieren stattdessen auf unadjustierten Analysen. Dieses Vorgehen kann zu weniger präzisen und damit weniger robusten Ergebnissen führen, insbesondere bei heterogener Anfallshäufigkeit zu Baseline.

2.5 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 7: Verzerrungspotential der Studie Marigold

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung Patientin/Patient	Verblindung Behandlungsperson	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
Marigold	Ja ¹⁾	Ja	Ja	Ja	Nein	Ja ²⁾	Unklar

¹⁾ Die Randomisierung erfolgte unstratifiziert und zentral mittels Interactive Web Response System (IWRS) im Verhältnis 1:1.

²⁾ Es bestehen relevante Baseline-Imbalancen im Ganaxolon-Arm (N = 50) gegenüber dem Placebo-Arm (N = 51) u. a. hinsichtlich der Anfallshäufigkeiten jeglicher Anfälle (Median: 116,0 vs. 60,0), der Anzahl anfallsfreier Tage jeglicher Anfälle (Median: 0 vs. 16,2) und schwerer motorischer Anfälle (11,9 vs. 23,8) sowie hinsichtlich des Anteils von nicht gehfähigen Personen (90,0 % vs. 74,5 %). Diese Imbalancen können auf eine höhere Krankheitslast im Ganaxolon-Arm im Vergleich zum Placebo-Arm zu Baseline hindeuten. Adjustierte Auswertungen wurden nicht vorgelegt.

Abkürzungen: IWRS: Interactive Web Response System.

Das Verzerrungspotential auf Studienebene wird als unklar eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 8 dargestellt.

Tabelle 8: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie Marigold

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Mortalität					
Todesfälle	Ja	Ja ¹⁾	Nein	Nein	Niedrig ²⁾
Morbidität					
Schwere motorische Anfälle ³⁾	Ja	Ja	Nein	Unklar ⁴⁾	Unklar ⁵⁾
Jegliche Anfälle	Ja	Ja ⁶⁾	Nein	Nein	Unklar ⁵⁾
Anfallsfreie Tage jeglicher Anfälle	Ja	Ja	Nein	Nein	Unklar ⁵⁾
CGI-I	Ja	Nein ⁷⁾	Nein	Nein	Hoch ⁵⁾
Lebensqualität					
QI-Disability	Ja	Nein ⁸⁾	Nein	Nein	Hoch ⁵⁾
Sicherheit					
Schwere UE (Gesamtrate, SOC/PT)	Ja	Ja ⁹⁾	Nein	Ja	Unklar ⁵⁾
SUE (Gesamtrate, SOC/PT)	Ja	Ja ⁹⁾	Nein	Ja	Unklar ⁵⁾
Abbruch wegen UE (Gesamtrate)	Ja	Ja ⁹⁾	Nein	Ja	Unklar ⁵⁾
UE, jeglichen Schweregrades (SOC/PT)	Ja	Ja ⁹⁾	Nein	Ja	Unklar ⁵⁾

¹⁾ Todesfälle werden im Rahmen der „Sicherheit“ erfasst. Die Sicherheitspopulation entspricht der ITT-Population.

²⁾ Es ist nicht davon auszugehen, dass die Imbalancen zu Studienbeginn einen Einfluss auf die Todesfälle während der Studie haben. Es traten keine Todesfälle im doppelblinden Studienzeitraum auf.

³⁾ Der primäre Endpunkt wird ergänzend dargestellt.

⁴⁾ Aufgrund fehlender spezifischer Angaben zu den Modalitäten der Anfallserhebung mittels Tagebuchs, zur Klassifikationsprüfung durch ESC und SIF/DRF sowie zu einer möglichen Schulung ist die korrekte Klassifikation der Anfälle als schwer bzw. primär oder sekundär nicht gänzlich nachvollziehbar. Die Auswirkungen auf das Verzerrungspotential sind unklar.

⁵⁾ Das Verzerrungspotential auf Studienebene beeinflusst das Verzerrungspotential auf Endpunktebene.

⁶⁾ Personen mit fehlenden Werten wurden von der Auswertung ausgeschlossen. Dies betrifft 1 Person (2,0 %) im Ganaxolon-Arm. Es wird demnach von keinem verzerrenden Effekt ausgegangen.

⁷⁾ Personen mit fehlenden Werten wurden von der Auswertung ausgeschlossen. Der Anteil fehlender Werte zu Woche 17 beträgt 14,0 % im Ganaxolon-Arm und 7,8 % im Placebo-Arm. Teilweise sind die fehlenden Werte jedoch auf 2 (4,0 %) Studienabbrechende im Ganaxolon Arm bzw. vollständig auf Studienabbrechende im Placebo-Arm 4 (7,8 %) zurückzuführen. Für die übrigen 5 (10,0 %) Personen des Ganaxolon-Arms sind die Gründe für die fehlenden Werte unklar. Das Verzerrungspotential wird aufgrund der fehlenden Werte als hoch angesehen.

⁸⁾ Personen mit fehlenden Werten wurden von der Auswertung ausgeschlossen. Der Anteil fehlender Werte zu Woche 17 beträgt 16,0 % im Ganaxolon-Arm und 15,7 % im Placebo-Arm. Teilweise sind die fehlenden Werte jedoch auf 2 (4,0 %) Studienabbrechende im Ganaxolon-Arm bzw. vollständig auf Studienabbrechende im Placebo-Arm 4 (7,8 %) zurückzuführen. Für die übrigen 6 (12,0 %) Personen des Ganaxolon-Arms und 4 (7,9 %) Personen des Placebo-Arms sind die Gründe für die fehlenden Werte unklar. Das Verzerrungspotential wird aufgrund der fehlenden Werte als hoch angesehen.

⁹⁾ Die Sicherheitspopulation entspricht der ITT-Population.

Abkürzungen: CGI-I: Clinical Global Impression of Improvement; DRF: Seizure Identification and Diagnostic Review Form; ESC: Epilepsie-Konsortium; ITT: Intention to treat; KI: Konfidenzintervall; PT: Preferred Term; QI-Disability: Quality of Life Inventory-Disability; SIF: Seizure Identification and Diagnostic Review Form; SOC: Systemorganklasse; (S)UE: (Schwerwiegendes) unerwünschtes Ereignis.

3 Ergebnisse der eingeschlossenen Studie

3.1 Studiencharakteristika und Studienmedikation

Allgemeine Angaben zur Studie Marigold finden sich in Tabelle 9. Insgesamt wurden 114 Personen gescreent, wovon 101 Personen in die Studie eingeschlossen wurden. 50 Personen wurden in die Interventionsgruppe (Ganaxolon) und 51 Personen in die Kontrollgruppe (Placebo) randomisiert.

Tabelle 9: Allgemeine Angaben; Studie Marigold

Studie Marigold Allgemeine Angaben	Ganaxolon N = 50	Placebo N = 51
ITT-/Sicherheitspopulation, n (%) ¹⁾	50 (100)	51 (100)
Beendigung der 17-wöchigen DBP, n (%)	48 (96,0)	47 (92,2)
Abbruch der Einnahme der Studienmedikation, n (%)	3 (6,0)	0 (0)
Aufgrund von:		
UE	2 (4,0) ²⁾	0 (0)
Patientenentscheidung (auch Eltern/Betreuende)	1 (2,0)	0 (0)
Abbruch der Studie vor Beendigung der 17-wöchigen DBP, n (%)	2 (4,0)	4 (7,8)
Aufgrund von:		
UE	1 (2,0)	4 (7,8)
Patientenentscheidung (auch Eltern/Betreuende)	1 (2,0)	0 (0)
Übergang in die OLE ³⁾ , n (%)	43 (86,0)	45 (88,2)
Mediane Behandlungsdauer Wochen (min; max)	16,9 (2,7; 26,1)	16,9 (0,3; 27,1)
Mediane Beobachtungsdauer Wochen (min; max)	k. A.	k. A.

¹⁾ Definition, siehe Kapitel 2.4. In der vorliegenden Studie stimmen ITT- und Sicherheitspopulation überein.

²⁾ Von diesen 2 Personen, die die Einnahme der Studienmedikation aufgrund eines UE abgebrochen haben, beendete 1 Person die Studie. Die Abbruchgründe der Studienmedikation entsprechen denen der Studienabbrüche.

³⁾ Gegenstand der vorliegenden Nutzenbewertung ist die 17-wöchige doppelblinde, placebokontrollierte Studienphase.

Abkürzungen: DBP: doppelblinde, placebokontrollierte Studienphase; ITT: Intention to treat; k. A.: keine Angaben; OLE: offene einarmige Studienphase (Open-Label Extension); UE: Unerwünschtes Ereignis.

Die Baseline-Charakteristika der eingeschlossenen Patientinnen und Patienten sind nachfolgend in Tabelle 10 zusammengefasst.

Tabelle 10: Charakterisierung der Studienpopulation der Studie Marigold

Studie Marigold Charakterisierung der Studienpopulation	Ganaxolon N = 50	Placebo N = 51
<i>Alter (Jahre)</i>		
Mittelwert (SD)	6,78 (4,71)	7,73 (4,38)
Median (min; max)	5,0 (2,0; 19,0 ¹⁾)	7,0 (2,0; 19,0 ¹⁾)
<i>Geschlecht, n (%)</i>		
männlich	11 (22,0)	10 (19,6)
weiblich	39 (78,0)	41 (80,4)
<i>Abstammung, n (%)</i>		
weiß	46 (92,0)	47 (92,2)
asiatisch	2 (4,0)	3 (5,9)
andere	2 (4,0)	1 (2,0)

Studie Marigold Charakterisierung der Studienpopulation	Ganaxolon N = 50	Placebo N = 51
<i>Studienland, n (%)</i>		
Australien	4 (8,0)	2 (3,9)
Frankreich	3 (6,0)	3 (5,9)
Israel	0 (0)	1 (2,0)
Italien	9 (18,0)	6 (11,8)
Polen	5 (10,0)	5 (9,8)
Russland	7 (14,0)	7 (13,7)
Vereinigtes Königreich	4 (8,0)	3 (5,9)
USA	18 (36,0)	24 (47,1)
<i>Region, n (%)</i>	k. A.	k. A.
EMEA		
Nordamerika		
Lateinamerika		
Asien/Pazifik		
<i>Körpergröße (cm)</i>		
n (%)	50 (100)	51 (100)
Mittelwert (SD)	112,6 (20,8)	119,6 (19,6)
Median (min; max)	109,0 (48,5; 160,0)	119,0 (82,5; 160,0)
<i>Körpergewicht (kg)</i>		
n (%)	50 (100)	51 (100)
Mittelwert (SD)	20,0 (8,90)	22,5 (9,05)
Median (min; max)	16,8 (9,7; 42,0)	19,1 (8,7; 46,5)
<i>Tanner-Stadium, n (%)</i>		
Stadium 1	31 (62,0)	30 (58,8)
Stadium 2	5 (10,0)	5 (9,8)
Stadium 3	1 (2,0)	8 (15,7)
Stadium 4	2 (4,0)	2 (3,9)
Stadium 5	3 (6,0)	1 (2,0)
Keine Angabe verfügbar	8 (16,0)	5 (9,8)
<i>Anzahl der Antikonvulsiva (Vorthherapie)</i>		
n (%)	48 (96,0)	50 (98,0)
Mittelwert (SD)	7,75 (3,53)	6,82 (3,30)
Median (min; max)	7,0 (2,0; 16,0)	7,0 (1,0; 14,0)
<i>Nicht antikonvulsive Wirkstoffe (Vorthherapie), n (%)</i>		
Ja	7 (14,0)	14 (27,5)
Nein	43 (86,0)	37 (72,5)
<i>Anfallstagebuch-Compliance zu Baseline²⁾</i>		
n (%)	49 (98,0)	51 (100)
Mittelwert (SD)	97,6 (5,79)	94,2 (12,8)
Median (min; max)	100,0 (73,8; 100)	100,0 (28,6; 100)
<i>Anfallshäufigkeiten für jegliche Anfälle³⁾ zu Baseline⁴⁾</i>		
n (%)	49 (98,0)	51 (100)
Mittelwert (SD)	178,8 (174,0)	141,1 (200,9)
Median (min; max)	116,0 (18,4; 651,3)	60,0 (7,0; 1021,3)
<i>Anfallsfreie Tage für jegliche Anfälle³⁾ zu Baseline⁴⁾</i>		
n (%)	49 (98,0)	51 (100)
Mittelwert (SD)	12,0 (19,0)	21,6 (22,9)
Median (min; max)	0 (0; 75,6)	16,2 (0; 81,0)

Studie Marigold Charakterisierung der Studienpopulation	Ganaxolon N = 50	Placebo N = 51
<i>Anfallshäufigkeiten für schwere motorische⁵⁾ Anfälle zu Baseline⁴⁾</i> n (%) Mittelwert (SD) Median (min; max)	49 (98,0) 115,4 (138,4) 54,0 (31,3; 147,3)	51 (100) 103,9 (173,0) 49,2 (0,7; 1021,3)
<i>Anfallsfreie Tage für schwere motorische Anfälle zu Baseline⁴⁾</i> n (%) Mittelwert (SD) Median (min; max)	49 (98,0) 22,6 (25,8) 11,9 (0; 82,9)	51 (100) 30,3 (27,1) 23,8 (0; 97,5)
<i>Alter beim ersten Anfall (Monate)</i> n (%) Mittelwert (SD) Median (min; max)	50 (100) 2,8 (2,70) 2,0 (0; 14)	51 (100) 2,3 (2,34) 2,0 (0; 10)
<i>Beginn der Anfälle bis zum 1. Lebensjahr, n (%)</i> Ja Nein	49 (98,0) 1 (2,0)	51 (100) 0 (0,0)
<i>Gehfähigkeit, n (%)</i> Ja Nein	5 (10,0) 45 (90,0)	13 (25,5) 38 (74,5)
<i>Alter bei Erreichen der Gehfähigkeit (sofern gehfähig), in Monaten</i> n (%) Mittelwert (SD) Median (min; max)	5 (10,0) 50,4 (32,2) 36,0 (24; 96)	13 (25,5) 45,5 (33,9) 33,0 (18; 120)
<i>Fehlende Fähigkeiten zur selbstständigen Fortbewegung bis zum 2. Lebensjahr, n (%)</i> Ja Nein	50 (100) 0 (0,0)	51 (100) 0 (0,0)
<i>Sprachfähigkeiten, n (%)</i> Patient kann Phrasen von 2-3 Wörtern sprechen Patient ist nicht sprachfähig Patient spricht einzelne Wörter	1 (2,0) 47 (94,0) 2 (4,0)	0 (0,0) 49 (96,1) 2 (3,9)
<i>Stereotypien, n (%)</i> Ja Nein	33 (66,0) 17 (34,0)	34 (66,7) 17 (33,3)
<i>Baseline QI-Disability (Gesamtscore)^{6) 7)}</i> n (%) Mittelwert (SD) Median (min; max)	49 (98,0) 53,0 (15,4) 51,6 (29,7; 91,4)	49 (96,1) 53,3 (15,7) 53,9 (15,6; 89,1)

¹⁾ Es wurden zwei 19-jährige erwachsene Patientinnen eingeschlossen und randomisiert jeweils einem Studienarm zugeordnet.

²⁾ Die Tagebuch-Compliance entspricht dem prozentualen Anteil der Tage mit verfügbaren Anfallsdaten an der Gesamtzahl der vorgesehenen Tage im jeweiligen Beobachtungsintervall (6-wöchige Baseline bzw. 17-wöchige DBP). Sie wird berechnet aus der Anzahl der Tage mit verfügbaren Anfallsdaten geteilt durch die Anzahl der vorgesehenen Tage, multipliziert mit 100.

³⁾ Jegliche Anfälle umfassen sämtliche aufgetretene Anfallsarten.

⁴⁾ Als Baseline-Phase (Teil der Screening-Phase) wurden die 6 Wochen vor Beginn der Behandlung mit der Studienmedikation definiert (frühestens 42 Tage und spätestens 48 Tage nach Visite 1).

⁵⁾ Schwere motorische Anfälle umfassen bilateral tonische Anfälle (anhaltende motorische Aktivität ≥ 3 Sekunden), generalisierte tonisch-klonische Anfälle, bilateral klonische Anfälle, atonische Anfälle/Sturzanfälle und fokal zu bilateral tonisch-klonische Anfälle.

⁶⁾ Baseline entspricht dem Zeitpunkt der Studienvsiste 2.

⁷⁾ Die Skalenspannweite beträgt 0–100. Höhere Werte zeigen eine bessere Lebensqualität an.

Abkürzungen: DBP: doppelblinde, placebokontrollierte Studienphase; QI-Disability: Quality of Life Inventory-Disability; SD: Standardabweichung.

Protokollverletzungen

Anhand der in der Studie dokumentierten Protokollverletzungen ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

Begleitmedikation

Gemäß den Einschlusskriterien mussten die Studienteilnehmenden während der gesamten Studie eine stabile Begleitmedikation mit ≥ 1 antikonvulsive Arzneimittel aufweisen. Eine Zusammenfassung zu den antikonvulsiven Begleitmedikamenten in der 17-wöchigen DBP ist nachfolgend abgebildet.

Tabelle 11: Begleitmedikation; Studie Marigold

Studie Marigold Antikonvulsive Begleitmedikation	Ganaxolon N = 50	Placebo N = 51
Personen mit ≥ 1 antikonvulsive Begleitmedikation, n (%)	49 (98,0)	48 (94,1)
Anzahl der Antikonvulsiva (Begleittherapie) n (%)	50 (100)	51 (100)
Mittelwert (SD)	2,6 (1,39)	2,2 (1,14)
Median (min; max)	2,0 (0; 6,0)	2,0 (0; 5,0)
Häufigste Antikonvulsiva (≥ 15 %), n (%)		
Valproat	18 (36,0)	16 (31,4)
Levetiracetam	13 (26,0)	13 (25,5)
Clobazam	12 (24,0)	13 (25,5)
Vigabatrin	10 (20,0)	12 (23,5)
Personen mit ≥ 1 Rescue-Medikation, n (%)	25 (50,0)	22 (43,1)

Abkürzungen: SD: Standardabweichung.

Ein Großteil der Patientinnen und Patienten (84 % bzw. 92 %) nahm zudem weitere Begleitmedikamente außer den anfallshemmenden Arzneimitteln ein. Zu den häufigsten (≥ 15 %) Begleitmedikationen und -therapien gehörten Paracetamol (31 %) und Macrofol 3350 (16 %).

Im Median erhielten 25 Personen (50 %) im Ganaxolon- und 22 Personen (43 %) im Placebo-Arm mindestens 1 Dosis einer Rescue-Medikation. Im Median wurden 8 bzw. 6 Dosen der Notfallmedikation in den jeweiligen Behandlungsarmen verabreicht.

Anhand der in der Studie Marigold dokumentierten Begleitmedikation ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

3.2 Mortalität

Es traten keine Todesfälle in der Studie Marigold bis Woche 17 auf.

3.3 Morbidität

Epileptische Anfälle

Die Ergebnisse zum Endpunkt „jegliche Anfälle“ werden in Tabelle 12 bis Tabelle 14 dargestellt. Der primäre Endpunkt „schwere motorische Anfälle“ wird ergänzend in Tabelle 15 dargestellt. Es zeigte sich in der ergänzenden Ergebnisdarstellung des primären Endpunktes zur medianen prozentualen Veränderung der Häufigkeit schwerer motorischer Anfälle gegenüber Baseline ein statistisch signifikanter Vorteil von Ganaxolon gegenüber Placebo ($p = 0,004$). In weiteren von der EMA geforderten Analysen zum primären Endpunkt zeigten die nachgereichten Analysen mittels negativ-binomialen Regressionsmodells kein statistisch signifikantes Ergebnis.

In Bezug auf den Endpunkt „jegliche Anfälle“ wurden über sämtliche Analysen hinweg keine statistisch signifikanten Ergebnisse berichtet.

Jegliche Anfälle

Tabelle 12: Veränderung der Häufigkeit jeglicher Anfälle der 17-wöchigen DBP gegenüber Baseline; Studie Marigold, ITT-Population (Datenschnitt: 19.05.2021)

Studie Marigold Häufigkeit¹⁾ jeglicher Anfälle²⁾	Ganaxolon N = 50	Placebo N = 51
<i>Baseline-Phase³⁾</i> n (%) Median [95%-KI] ⁴⁾	49 (98,0) 116,0 [75,3; 173,3]	51 (100) 60,0 [46,0; 120,0]
<i>17-wöchige DB-Behandlungsperiode⁵⁾</i> n (%) Median [95%-KI] ⁴⁾	50 (100) 79,2 [56,9; 132,8]	51 (100) 72,8 [51,2; 100,2]
<i>Prozentuale Veränderung ggü. Baseline⁶⁾</i> n (%) Median [95%-KI] ⁴⁾	49 (98,0) -19,1 [-36,8; -3,28]	51 (100) -8,91 [-20,5; 9,60]
Hodges-Lehmann-Schätzer [95%-KI] ⁴⁾ ; p-Wert ⁷⁾	-17,4 [-36,4; 0,33]; 0,051	

¹⁾ Angaben basieren auf der Summe einzelner Anfälle, zählbarer Anfälle und Cluster mit nicht zählbaren Anfällen (jedes Cluster mit nicht zählbaren Anfällen zählt als 1 Anfall). Die Häufigkeit der Anfälle ist normiert auf 28 Tage.

²⁾ Jegliche Anfälle umfassen sämtliche aufgetretene Anfallsarten.

³⁾ Die Baseline-Phase umfasst die 6 Wochen vor Beginn der Behandlung mit der Studienmedikation (frühestens 42 Tage und spätestens 48 Tage nach Visite 1).

⁴⁾ Verteilungsfreies 95%-KI des Medians.

⁵⁾ Die 17-wöchige Behandlungsphase umfasst den Tag nach der ersten Gabe der Studienmedikation bis zum Tag vor Visite 5 (Woche 17), falls verfügbar; andernfalls bis zum letzten Tag mit Anfallsdaten. Schließt eine Person die DBP ohne Visite 5 ab, jedoch mit Taper-Visite und geht nicht in die OLE über, endet das Intervall am Tag vor der Taper-Visite.

⁶⁾ Die Berechnung der prozentualen Änderung gegenüber Baseline erfolgte nur für Personen deren Baselinewert nicht null war.

⁷⁾ p-Wert berechnet mit dem Wilcoxon-Rangsummen-Test.

Abkürzungen: DBP: doppelblinde, placebokontrollierte Studienphase; ITT: Intention to treat; KI: Konfidenzintervall; OLE: offene einarmige Studienphase (Open-Label Extension).

Tabelle 13: Responderanalysen zur Reduktion der Häufigkeit jeglicher Anfälle um $\geq 25\%$, $\geq 50\%$ und $\geq 75\%$ zu Woche 17; ITT-Population (Datenschnitt: 19.05.2021)

Studie Marigold Häufigkeit ¹⁾ jeglicher Anfälle ²⁾	Ganaxolon N = 50	Placebo N = 51
≥ 25 % Reduktion		
Verfügbare Werte zu Woche 17, n (%)	49 (98,0) ³⁾	51 (100) ³⁾
Responder, n (%)	22 (44,9)	15 (29,4)
RR [95%-KI] ⁴⁾ ; p-Wert ⁵⁾	1,63 [0,98; 2,72]; 0,07	
≥ 50 % Reduktion		
Verfügbare Werte zu Woche 17, n (%)	49 (98,0) ³⁾	51 (100)
Responder, n (%)	10 (20,4)	4 (7,8)
RR [95%-KI] ⁴⁾ ; p-Wert ⁵⁾	2,60 [0,87; 7,75]; 0,09	
≥ 75 % Reduktion		
Verfügbare Werte zu Woche 17, n (%)	49 (98,0) ³⁾	51 (100)
Responder, n (%)	5 (10,2)	1 (2,0)
RR [95%-KI] ⁴⁾ ; p-Wert ⁵⁾	5,20 [0,63; 43,0]; 0,11	

¹⁾ Angaben basieren auf der Summe einzelner Anfälle, zählbarer Anfälle und Cluster mit nicht zählbaren Anfällen (jedes Cluster mit nicht zählbaren Anfällen zählt als 1 Anfall).

²⁾ Jegliche Anfälle umfassen sämtliche aufgetretene Anfallsarten.

³⁾ Fehlende Werte gehen nicht in die Analyse ein.

⁴⁾ KIs berechnet mit der Wald-Methode.

⁵⁾ p-Wert berechnet mit dem exakten Fisher-Test.

Abkürzungen: ITT: Intention to treat; KI: Konfidenzintervall; RR: Relatives Risiko.

Tabelle 14: Veränderung der anfallsfreien Tage jeglicher Anfälle der 17-wöchigen DBP gegenüber Baseline; Studie Marigold, ITT-Population (Datenschnitt: 19.05.2021)

Studie Marigold Anfallsfreie Tage ¹⁾ jeglicher Anfälle ²⁾	Ganaxolon N = 50	Placebo N = 51
Baseline-Phase³⁾		
n (%)	49 (98,0)	51 (100)
Median [95%-KI] ⁴⁾	0,00 [0; 9,52]	16,2 [9,52; 23,8]
17-wöchige DB-Behandlungsperiode⁵⁾		
n (%)	50 (100)	51 (100)
Median [95%-KI] ⁴⁾	3,80 [0,87; 15,8]	20,5 [7,14; 36,8]
Prozentuale Veränderung Woche 17 zu Baseline⁶⁾		
n (%)	49 (98,0)	51 (100)
Median [95%-KI] ⁴⁾	0,07 [0; 2,51]	0,00 [0; 9,35]
Hodges-Lehmann-Schätzer [95%-KI] ⁴⁾ ; p-Wert ⁷⁾	0,00 [-4,32; 2,56]; 0,78	

¹⁾ Der Anteil anfallsfreier Tage entspricht der Anzahl der Tage ohne Anfälle geteilt durch die Gesamtanzahl der Tage mit Anfallsdaten in den jeweiligen Studienphasen (DBP, Baseline), multipliziert mit 100.

²⁾ Jegliche Anfälle umfassen sämtliche aufgetretene Anfallsarten.

³⁾ Die Baseline-Phase umfasst die 6 Wochen vor Beginn der Behandlung mit der Studienmedikation (frühestens 42 Tage und spätestens 48 Tage nach Visite 1).

⁴⁾ Verteilungsfreies 95%-KI des Medians.

⁵⁾ Die 17-wöchige Behandlungsphase umfasst den Tag nach der ersten Gabe der Studienmedikation bis zum Tag vor Visite 5 (Woche 17), falls verfügbar; andernfalls bis zum letzten Tag mit Anfallsdaten. Schließt eine Person

die DBP ohne Visite 5 ab, jedoch mit Taper-Visite und geht nicht in die OLE über, endet das Intervall am Tag vor der Taper-Visite.

⁶⁾ Die Berechnung der prozentualen Änderung gegenüber Baseline erfolgte nur für Personen, deren Baselinewert nicht null war.

⁷⁾ p-Wert berechnet mit dem Wilcoxon-Rangsummen-Test.

Abkürzungen: DBP: doppelblinde, placebokontrollierte Studienphase; ITT: Intention to treat; KI: Konfidenzintervall; OLE: offene einarmige Studienphase (Open-Label Extension).

Schwere motorische Anfälle (ergänzend dargestellt)

Tabelle 15: Veränderung schwerer motorischer Anfälle der 17-wöchigen DBP gegenüber Baseline; Studie Marigold, ITT-Population (Datenschnitt: 19.05.2021) (ergänzend dargestellt)

Studie Marigold Häufigkeit¹⁾ schwerer motorischer Anfälle²⁾³⁾ (ergänzend dargestellt)	Ganaxolon N = 50	Placebo N = 51
<i>Baseline-Phase⁴⁾</i> n (%) Median [95%-KI] ⁵⁾	49 (98,0) 54,0 [38,2; 106,7]	51 (100) 49,2 [32,2; 60,7]
<i>17-wöchige DB-Behandlungsperiode⁶⁾</i> n (%) Median [95%-KI] ⁵⁾	50 (100) 45,0 [31,8; 76,0]	51 (100) 55,5 [35,8; 80,1]
<i>Prozentuale Veränderung ggü. Baseline⁷⁾</i> n (%) Median [95%-KI] ⁵⁾	49 (98,0) -30,7 [-35,9; -12,0]	51 (100) -6,90 [-16,5; 15,3]
Hodges-Lehmann-Schätzer [95%-KI] ⁵⁾ ; p-Wert ⁸⁾	-27,1 [-47,9; -9,55]; 0,004	

¹⁾ Angaben basieren auf der Summe einzelner Anfälle, zählbarer Anfälle und Cluster mit nicht zählbaren Anfällen (jedes Cluster mit nicht zählbaren Anfällen zählt als 1 Anfall). Die Häufigkeit der Anfälle ist normiert auf 28 Tage.

²⁾ Schwere motorische Anfälle umfassen bilateral tonische (≥ 3 s), generalisiert tonisch-klonische, atonische/ Sturzanfälle, bilateral klonische und fokal zu bilateral tonisch-klonische Anfälle.

³⁾ Primärer Endpunkt.

⁴⁾ Die Baseline-Phase umfasst die 6 Wochen vor Beginn der Behandlung mit der Studienmedikation (frühestens 42 Tage und spätestens 48 Tage nach Visite 1).

⁵⁾ Verteilungsfreies 95%-KI des Medians.

⁶⁾ Die 17-wöchige Behandlungsphase umfasst den Tag nach der ersten Gabe der Studienmedikation bis zum Tag vor Visite 5 (Woche 17), falls verfügbar; andernfalls bis zum letzten Tag mit Anfallsdaten. Schließt eine Person die DBP ohne Visite 5 ab, jedoch mit Taper-Visite und geht nicht in die OLE über, endet das Intervall am Tag vor der Taper-Visite.

⁷⁾ Die Berechnung der prozentualen Änderung gegenüber Baseline erfolgte nur für Personen, deren Baselinewert nicht null war.

⁸⁾ p-Wert berechnet mit dem Wilcoxon-Rangsummen-Test.

Abkürzungen: DBP: doppelblinde, placebokontrollierte Studienphase; ITT: Intention to treat; KI: Konfidenzintervall; OLE: offene einarmige Studienphase (Open-Label Extension).

Globaler Betreuerereindruck (CGI-I)

Die Ergebnisse zur Veränderung im globalen klinischen Eindruck der Eltern/Betreuenden sind nachfolgend in Tabelle 18 dargestellt. Die Ergebnisse der Responderanalysen zu Woche 17 gegenüber Baseline sind statistisch nicht signifikant ($p = 0,10$). Die beiden durchgeführten Sensitivitätsanalysen mit Imputation fehlender Werte bestätigen das Ergebnis ($p = 0,23$ in beiden Analysen). Die Ergebnisse des CGI-I der Klinikerinnen und Kliniker stimmen mit der Bewertung durch die Eltern/Betreuenden zu Woche 17 überein. Es zeigen sich keine signifikanten Gruppenunterschiede ($p = 0,35$).

Tabelle 16: Anteil der Personen mit einer Verbesserung¹⁾ des globalen klinischen Betreuerindrucks der Verbesserung (CGI-I–Eltern/Betreuende) zu Woche 17 gegenüber Baseline (post hoc); ITT-Population (Datenschnitt: 19.05.2021)

Studie Marigold CGI-I (Eltern/Betreuende)	Ganaxolon N = 50	Placebo N = 51
Verfügbare Werte zu Woche 17, n (%)	43 (86,0) ²⁾	47 (92,2) ²⁾
Responder, n (%)	27 (62,8)	21 (44,7)
RR [95%-KI] ³⁾ ; p-Wert ⁴⁾	1,41 [0,95; 2,08]; 0,10	

¹⁾ Jegliche Verbesserung war definiert als die Summe aus „geringfügig verbessert“, „stark verbessert“ und „sehr stark verbessert“.

²⁾ Fehlende Werte gehen nicht in die Analyse ein. Der Anteil fehlender Werte beträgt 14,0 % im Ganaxolon-Arm und 7,8 % im Placebo-Arm. Die Gründe für das Fehlen sind nicht berichtet. Teilweise sind die fehlenden Werte jedoch auf 2 (4,0 %) Studienabbrechende im Ganaxolon-Arm bzw. 4 (7,8 %) Personen im Placebo-Arm zurückzuführen. Es fand keine Erhebung zu Baseline statt.

³⁾ KIs berechnet mit der Wald-Methode.

⁴⁾ p-Wert berechnet mit dem exakten Fisher-Test.

Abkürzungen: CGI-I: Clinical Global Impression of Improvement; ITT: Intention to treat; KI: Konfidenzintervall; RR: Relatives Risiko.

3.4 Lebensqualität

Quality of Life Inventory-Disability (QI-Disability)

Die Ergebnisse zum Anteil der Personen mit einer Verbesserung um 15 % der Skalenspannweite sind nachfolgend für alle Skalen des QI-Disability in Tabelle 17 abgebildet. Im Gesamtscore des QI-Disability zeigten sich keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen ($p = 0,71$). Die Sensitivitätsanalysen zur Berücksichtigung fehlender Werte bestätigten diese Ergebnisse ($p = 0,36$ bzw. $0,91$). Auffällig ist, dass die Anzahl der Responder im Gesamtscore gegenüber jenen in den Domänenscores deutlich geringer ausfällt.

Tabelle 17: Lebensqualität gemäß dem QI-Disability-Fragebogen, Anzahl der Personen mit einer Verbesserung des QI-Disability um mindestens 15 % der Skalenspannweite zu Woche 17 (post hoc), ITT-Population (Datenschnitt: 19.05.2021)

Studie Marigold QI-Disability ¹⁾	Ganaxolon N = 50	Placebo N = 51
Domäne „Körperliche Gesundheit“ (ergänzend)		
Verfügbare Werte zu Woche 17, n (%)	44 (88,0) ²⁾	46 (90,2) ²⁾
Responder, n (%)	18 (40,9)	20 (43,5)
RR [95%-KI] ³⁾	0,94 [0,58; 1,52]	
Domäne „Positive Gefühle“ (ergänzend)		
Verfügbare Werte zu Woche 17, n (%)	42 (84,0) ²⁾	46 (90,2) ²⁾
Responder, n (%)	18 (42,9)	15 (32,6)
RR [95%-KI] ³⁾	1,30 [0,77; 2,22]	
Domäne „Negative Emotionen“ (ergänzend)		
Verfügbare Werte zu Woche 17, n (%)	43 (86,0) ²⁾	45 (88,2) ²⁾
Responder, n (%)	17 (39,5)	22 (48,9)
RR [95%-KI] ³⁾	0,81 [0,51; 1,30]	

Studie Marigold QI-Disability ¹⁾	Ganaxolon N = 50	Placebo N = 51
Domäne „Soziale Interaktion“ (ergänzend)		
Verfügbare Werte zu Woche 17, n (%)	43 (86,0) ²⁾	44 (86,3) ²⁾
Responder, n (%)	18 (41,9)	15 (34,1)
RR [95%-KI] ³⁾	1,22 [0,72; 2,08]	
Domäne „Freizeit und Außenbereich“ (ergänzend)		
Verfügbare Werte zu Woche 17, n (%)	43 (86,0) ²⁾	44 (86,3) ²⁾
Responder, n (%)	21 (48,8)	14 (31,8)
RR [95%-KI] ³⁾	1,52 [0,90; 2,55]	
Domäne „Selbstständigkeit“ (ergänzend)		
Verfügbare Werte zu Woche 17, n (%)	43 (86,0)	45 (88,2)
Responder, n (%)	24 (55,8)	14 (31,1)
RR [95%-KI] ³⁾	1,77 [1,07; 2,92]	
Gesamtscore		
Verfügbare Werte zu Woche 17, n (%)	42 (84,0) ²⁾	43 (84,3) ²⁾
Responder, n (%)	4 (9,5)	3 (7,0)
RR [95%-KI] ³⁾ ; p-Wert ⁴⁾	1,32 [0,35; 4,99]; 0,71	

¹⁾ Die Skalenspannweite beträgt 0–100. Höhere Werte zeigen eine bessere Lebensqualität an.

²⁾ Fehlende Werte gehen nicht in die Analyse ein. Die Anzahl der fehlenden Werte beträgt domänenspezifisch zwischen 12,0 %–16,0 % im Ganaxolon-Arm und 9,8 %–13,7 % im Placebo-Arm. Im Gesamtscore beträgt der Anteil fehlender Werte 16,0 % im Ganaxolon-Arm und 15,7 % im Placebo-Arm. Die Gründe für das Fehlen sind nicht berichtet; teilweise sind die fehlenden Werte jedoch auf 2 (4,0 %) Studienabbrechende im Ganaxolon-Arm bzw. 4 (7,8 %) Personen im Placebo-Arm zurückzuführen. Zu Baseline betrug der Anteil fehlender Werte in der Analyse (basierend auf dem Gesamtscore) 4,0 % bzw. 2,0 %.

³⁾ KIs berechnet mit der Wald-Methode.

⁴⁾ p-Wert berechnet mit dem exakten Fisher-Test.

Abkürzungen: ITT: Intention to treat; KI: Konfidenzintervall; QI-Disability: Quality of Life Inventory-Disability; RR: Relatives Risiko.

3.5 Sicherheit

Im Folgenden werden die Ergebnisse der Endpunktkategorie „Sicherheit“ der 17-wöchigen doppel-blinden, placebokontrollierten Studienphase dargestellt. Die mediane Behandlungszeit in beiden Behandlungsarmen war identisch und betrug 16,9 Wochen. Informationen zur medianen Beobachtungsdauer in beiden Behandlungsarmen liegen nicht vor.

Der pU legt post hoc zusätzliche Auswertungen unter Nichtberücksichtigung von Progressionsereignissen (epileptische Anfälle) für die zusammenfassenden UE der DBP vor, welche für die Nutzenbewertung herangezogen werden. Es ist jedoch unklar, inwiefern weitere nichtberücksichtigte Ereignisse der Grunderkrankung in die Auswertungen eingingen (siehe Kapitel 2.3.4).

Tabelle 18: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE unter Ausschluss krankheitsbezogener Ereignisse (Anfälle) während der 17-wöchigen DBP; Sicherheitspopulation, Studie Marigold (Datenschnitt: 19.05.2021)

Studie Marigold Zusammenfassung der UE	Ganaxolon N = 50 n (%)	Placebo N = 51 n (%)	RR ¹⁾ [95%-KI] ²⁾ ; p-Wert ³⁾
<i>Personen mit mindestens einem ...</i>			
UE (ergänzend dargestellt)	42 (84,0)	45 (88,2)	0,95 [0,81; 1,11]; 0,58
schweren UE ⁴⁾	1 (2,0)	1 (2,0)	1,02 [0,07; 15,9]; 1,00
SUE	6 (12,0)	4 (7,8)	1,53 [0,46; 5,10]; 0,53
UE, das zum Abbruch der Studienmedikation führte	1 (2,0)	2 (3,9)	0,51 [0,05; 5,45]; 1,00

¹⁾ RR post hoc berechnet.

²⁾ KIs post hoc berechnet mit der Wald-Methode.

³⁾ Berechnung der p-Werte erfolgte post hoc mit dem exakten Fisher-Test.

⁴⁾ Es wurden studieneigene Kriterien zur Schweregradeinteilung verwendet.

Abkürzungen: KI: Konfidenzintervall; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Unerwünschte Ereignisse

Die UE jeglichen Schweregrades mit einer Inzidenz $\geq 10\%$ sind nachfolgend abgebildet. Das PT Somnolenz trat im Ganaxolon-Arm statisch signifikant häufiger auf als im Vergleichsarm ($p = 0,02$). Effektschätzer auf Ebene der SOC wurden nicht berichtet. In Gesamtschau der Ergebnisse ist nicht davon auszugehen, dass sich auf Ebene der SOC ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen ergibt.

Tabelle 19: UE mit Inzidenz $\geq 10\%$ in der Studie Marigold; Sicherheitspopulation (Datenschnitt: 19.05.2021)

Studie Marigold UE MedDRA-Systemorganklasse Preferred Term	Ganaxolon N = 50 n (%)	Placebo N = 51 n (%)	RR ¹⁾ [95%-KI] ²⁾ ; p-Wert ³⁾
Erkrankungen des Nervensystems	27 (54,0)	20 (39,2)	k. A.
Somnolenz	18 (36,0)	8 (15,7)	2,30 [1,10; 4,79]; 0,02
Anfälle	7 (14,0)	9 (17,6)	0,79 [0,32; 1,97]; 0,79
Infektionen und parasitäre Erkrankungen	22 (44,0)	26 (51,0)	k. A.
Infektion der oberen Atemwege	5 (10,0)	3 (5,9)	1,70 [0,43; 6,74]; 0,49
Erkrankungen des Gastrointestinaltrakts	14 (28,0)	22 (43,1)	k. A.
Erbrechen	5 (10,0)	10 (19,6)	0,51 [0,19; 1,39]; 0,26
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	13 (26,0)	8 (15,7)	k. A.
Fieber	9 (18,0)	4 (7,8)	2,29 [0,76; 6,97]; 0,15
Erkrankungen der Atemwege, des Brustraums und Mediastinums	10 (20,0)	10 (19,6)	k. A.
Psychiatrische Erkrankungen	8 (16,0)	9 (17,6)	k. A.
Stoffwechsel- und Ernährungsstörungen	5 (10,0)	2 (3,9)	k. A.
Erkrankungen der Haut und des Unterhautgewebes	3 (6,0)	7 (13,7)	k. A.

¹⁾ RR post hoc berechnet.

²⁾ KIs post hoc berechnet mit der Wald-Methode.

³⁾ Berechnung der p-Werte erfolgte post hoc mit dem exakten Fisher-Test.

Abkürzungen: k. A.: keine Angabe; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Term; RR: Relatives Risiko; UE: Unerwünschtes Ereignis.

Es traten keine schweren UE bei $\geq 5\%$ der Patientinnen und Patienten in mindestens einem Behandlungsarm auf. Die aufgetretenen SUE $\geq 5\%$ sind in nachfolgender Tabelle berichtet.

Tabelle 20: SUE, die bei $\geq 5\%$ der Patientinnen und Patienten in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term in der Studie Marigold; Sicherheitspopulation (Datenschnitt: 19.05.2021)

Studie Marigold SUE MedDRA-Systemorganklasse Preferred Term	Ganaxolon N = 50 n (%)	Placebo N = 51 n (%)	RR [95%-KI]; p-Wert
Infektionen und parasitäre Erkrankungen¹⁾	3 (6,0)	2 (3,9)	k. A.

¹⁾ Laut Modul 4 sollten keine SUE $\geq 5\%$ aufgetreten sein. Die Angaben stammen aus dem CSR. Es wurden somit keine RR-, 95%-KI- und p-Werte berichtet.

Abkürzungen: CSR: Clinical Study Report; k. A.: keine Angabe; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Term; RR: Relatives Risiko; SUE: Schwerwiegendes unerwünschtes Ereignis.

4 Diskussion der Methodik und Ergebnisse

4.1 Design und Methodik der Studie

Studiendesign

Die Studie Marigold ist eine randomisierte, doppelblinde, Placebo-kontrollierte, multizentrische Studie im Parallelgruppendesign der Phase III zur Untersuchung der Wirksamkeit und Sicherheit von Ganaxolon gegenüber Placebo bei Kindern und Erwachsenen zwischen 2 und 21 Jahren mit CDD. Die Studie ist abgeschlossen.

Der primäre Endpunkt der Studie Marigold war die prozentuale Veränderung in der Häufigkeit der schweren motorischen Anfälle in der 17-wöchigen Behandlungsperiode. Als weitere Endpunkte wurden u. a. jegliche Anfälle, der globale Betreuereindruck mittels CGI und Lebensqualität mittels QI-Disability sowie die Sicherheit erhoben.

Die in 36 Studienzentren durchgeführte Studie besteht aus einer 14-wöchigen Screening-Periode, im Rahmen derer epileptische Anfälle mittels Anfallstagebücher erhoben wurden. Die Screening-Periode setzt sich wiederum aus einer 8-wöchigen historischen und einer 6-wöchigen prospektiven Baseline-Periode zusammen, während Letztere als Referenzzeitraum diente. Nach der Baseline-Phase wurden die Patientinnen und Patienten im Verhältnis 1:1 unstratifiziert den Behandlungsarmen Ganaxolon oder Placebo zugeteilt. Im Anschluss erfolgte eine 17-wöchige doppelblinde Behandlungsperiode (DBP), bestehend aus einer 4-wöchigen Titrations- und einer 13-wöchigen Erhaltungsphase, gefolgt von einer offenen, einarmigen Studienphase über 216 Wochen. Bei vorzeitigem Studienabbruch war sowohl während der doppelblinden, placebokontrollierten als auch während der offenen, einarmigen Studienphase (OLE) eine 2-wöchige Sicherheitsnachbeobachtung vorgesehen. Aufgrund der fehlenden Kontrollgruppe in der OLE basiert die vorliegende Nutzenbewertung ausschließlich auf der 17-wöchigen doppelblinden, placebokontrollierten Phase. Für die vorliegende Nutzenbewertung wird der präspezifizierte Datenschnitt zum Ende der DBP mit den aktualisierten Daten vom 19.05.2021 herangezogen. Die mediane Behandlungsdauer betrug 16,9 Wochen in beiden Behandlungsarmen.

Studienpopulation

Laut Einschlusskriterien mussten die Studienteilnehmenden eine gendiagnostisch bestätigte CDKL5-Variante, früh einsetzende, schwer kontrollierbare Anfälle sowie neurologische Entwicklungsstörungen aufweisen. Zusätzlich waren mindestens 16 schwere motorische Anfälle innerhalb von 28 Tagen in beiden 4-Wochen-Perioden der 8-wöchigen historischen Screening-Erhebung erforderlich, sowie eine unzureichende Anfallskontrolle unter mindestens zwei Antikonvulsiva.

Die Studienpopulation umfasst insgesamt 101 Patientinnen und Patienten, davon entfallen 50 Personen auf den Interventionsarm und 51 Personen auf den Kontrollarm. Von den randomisierten Patientinnen und Patienten haben 48 Personen (96 %) im Ganaxolon-Arm und 47 Personen (92 %) im Placebo-Arm die Studie beendet. Angaben zur medianen Beobachtungsdauer liegen nicht vor. Insgesamt haben 6 Personen (6 %), davon 2 (4 %) im Ganaxolon-Arm und 4 (8 %) im Placebo-Arm die Studie vor Ende der 17-wöchigen DBP abgebrochen.

Die Studienteilnehmenden stammten überwiegend aus den USA (36 % bzw. 47 %) und Europa (Frankreich, Italien, Polen, Vereinigtes Königreich zusammen: 34 % bzw. 35 %) sowie aus Russland (14 % bzw. 14 %), Australien (8 % bzw. 4 %) und Israel (0 % bzw. 2 %).

Das mediane Alter der Untersuchten war zwischen den Gruppen leicht unterschiedlich und betrug 5 bzw. 7 Jahre im Interventions- bzw. Placebo-Arm (min; max jeweils 2; 19 Jahre). Es wurden zudem deutlich mehr weibliche (78 % bzw. 80 %) als männliche (22 % bzw. 20 %) Betroffene in die Studie eingeschlossen, was jedoch mit der bekannten Geschlechterverteilung bei CDD übereinstimmt [20]. Die Verteilung der Tanner-Stadien ist zwischen den Behandlungsarmen überwiegend ähnlich, jedoch mit deutlich mehr Patientinnen und Patienten in Stadium 3 im Placebo-Arm (16 % vs. 2 %). Die Interpretation dieser Unterschiede ist jedoch aufgrund der geringen Fallzahlen in den jeweiligen Stadien nicht abschließend möglich.

Hinsichtlich der krankheitsspezifischen Baseline-Charakteristika (Tabelle 10) waren Imbalancen zwischen den Behandlungsgruppen zu verzeichnen, die auf eine höhere Krankheitslast im Ganaxolon-Arm hindeuten können. Die Anfallshäufigkeiten jeglicher Anfälle im Interventionsarm waren im Median deutlich höher als im Placebo-Arm (116 bzw. 60 Anfälle). Damit wiesen Patientinnen und Patienten aus dem Ganaxolon-Arm im Vergleich zum Placebo-Arm eine deutlich niedrigere mediane Anzahl anfallsfreier Tage sowohl für jegliche Anfälle (0 bzw. 16 Tage) als auch für schwere motorische Anfälle (12 bzw. 24 Tage) in der 6-wöchigen Baseline-Periode auf. Zudem waren im Interventionsarm 90 % der Personen nicht gehfähig, gegenüber 75 % im Placebo-Arm. Die Anfälle begannen in beiden Behandlungsarmen im Median im Alter von 2 Monaten.

Studienmedikation

In der Studie Marigold erfolgte nach einer 4-wöchigen Aufdosierung dreimal täglich eine gewichtsabhängige Gabe der Erhaltungsdosis Ganaxolon in Höhe von 63 mg/kg für Patientinnen und Patienten ≤ 28 kg und 1.800 mg für alle Personen > 28 kg. Die Ganaxolon-Gabe war fachinformationskonform; die Placebo-Verabreichung erfolgte wie bei der Intervention, nur ohne den aktiven Wirkstoff.

Alle randomisierten Patientinnen und Patienten hatten mindestens eine Dosis der Studienmedikation erhalten. Damit entspricht die Sicherheitspopulation allen randomisierten Personen (ITT-Population). Die mediane Behandlungsdauer betrug in beiden Studienarmen 16,9 Wochen. Die Studienmedikation wurde von 3 Personen (6 %) ausschließlich im Ganaxolon-Arm abgebrochen, eine Person beendete die Studie, 2 Personen brachen aufgrund eines UE die Behandlung ab.

In Hinblick auf die Begleitmedikationen war für den überwiegenden Anteil der Patientinnen und Patienten in der Studie Marigold eine vorherige Anwendung antikonvulsiver Medikation (98 % bzw. 94 %) vorherrschend. Falls bestehend, sollte das vorherige, stabile Behandlungsschema für CDD mit bis zu 4 anfallshemmenden Medikamenten, außer bei Unverträglichkeit oder Kontraindikation, über die gesamte Studiendauer beibehalten werden. Die Anzahl und Art der eingesetzten Antiepileptika war zwischen den Behandlungsarmen vergleichbar (Tabelle 11).

4.2 Zulassungsstatus und Zulassungspopulation

Zulassungsstatus und Anwendungsgebiet

Ganaxolon (Ztalmy) ist zugelassen als Zusatzbehandlung von epileptischen Anfällen im Zusammenhang mit einer Cyclin-abhängigen Kinase-ähnlichen 5 (CDKL5)-Mangelkrankung (CDD) bei Patienten im Alter von 2 bis 17 Jahren und kann darüber hinaus bei Personen ab 18 Jahren weiter angewendet werden. Gemäß Fachinformation beträgt die nach einer Aufdosierungsphase zugelassene Tageshöchst-dosis Ganaxolon 63 mg/kg/Tag für Patientinnen und Patienten ≤ 28 kg und 1.800 mg/Tag für alle Personen > 28 kg, verabreicht als orale Lösung [9].

Das Anwendungsgebiet scheint weitestgehend durch die in die Studie Marigold eingeschlossenen Patientinnen und Patienten abgedeckt. Zu berücksichtigen ist, dass abweichend des Altersbereiches in der Fachinformation (2–17 Jahre) auch junge Erwachsene bis 21 Jahre in der Studie Marigold (2–21 Jahre) eingeschlossen werden konnten. Dies betrifft in der Marigold-Studie 2 Personen (2 %) im Alter von jeweils 19 Jahren. Diese Abweichung ist als vertretbar einzustufen, da auch die Fachinformation eine Weiterbehandlung Erwachsener ab 18 Jahren bei entsprechender Indikation vorsieht (jedoch einen Behandlungsbeginn bereits erwachsener Personen explizit ausschließt) und zudem der Anteil erwachsener Personen in der Studie Marigold sehr gering ist.

Ganaxolon ist laut Fachinformation als Zusatztherapie zu anderen Antikonvulsiva vorgesehen. Entsprechend der Einschlusskriterien der Marigold-Studie konnten hingegen auch Personen eingeschlossen werden, welche keine begleitende antiepileptische Therapie erhielten. Dies betraf insgesamt 4 (4 %) Personen (1 Person im Ganaxolon- und 3 Personen im Placebo-Arm; Tabelle 11). Die Abweichung von der Zulassungspopulation wird aufgrund des geringen Anteils von Personen ohne Zusatztherapie in der Studienpopulation als vertretbar angesehen.

Übertragbarkeit auf deutschen Versorgungskontext

Bei der Studie Marigold handelt es sich um eine multizentrische Studie, bei der 42 % der Personen im Ganaxolon-Arm und 33 % der Personen im Placebo-Arm an Studienzentren in Europa behandelt wurden, jedoch an keinem Studienzentrum in Deutschland. Der Großteil der eingeschlossenen Patientinnen und Patienten wurde in den USA (36 % bzw. 47 %) behandelt. Bis auf Russland fand die Behandlung eher in westlichen Nationen (Australien, Israel) statt. Insgesamt wird von einem vergleichbaren Versorgungskontext ausgegangen.

4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit

Das Verzerrungspotential auf Studienebene wird als unklar eingeschätzt. Hauptsächlich basiert das unklare Verzerrungspotential auf den beobachteten Baseline-Imbalancen in der Anfallshäufigkeit zwischen den Behandlungsarmen, welche sich in einer erhöhten Krankheitslast des Ganaxolon-Arms äußern können.

In Bezug auf die Ergebnissicherheit und externe Validität ist der kurze doppelblinde Beobachtungszeitraum von 17 Wochen einschränkend zu berücksichtigen. Aussagen zu längerfristigen Veränderungen hinsichtlich Wirksamkeit und Sicherheit sind daher nicht möglich.

Im Hinblick auf die Baseline-Imbalancen, welche sich primär in einer höheren Anfallshäufigkeit im Ganaxolon-Arm äußern, ist anzumerken, dass weder eine stratifizierte Randomisierung durchgeführt wurde noch eine Adjustierung der Analysen hinsichtlich der Anfallshäufigkeiten zu Baseline erfolgte. Dieses Vorgehen kann zu weniger präzisen bzw. robusten Ergebnissen führen. Dies ist insbesondere relevant, wenn die Anfallshäufigkeiten zwischen den Patientinnen und Patienten heterogen sind. Auf das Fehlen adjustierter Analysen in Form von parametrischen Tests hat auch die Europäische Arzneimittelagentur (EMA) hingewiesen.

4.4 Mortalität

Todesfälle wurden im Rahmen der Sicherheit erfasst. Im Verlauf der 17-wöchigen randomisierten Behandlungsphase der Studie Marigold ist keine Person verstorben. Aufgrund der sehr kurzen Beobachtungszeit von 17 Wochen ist keine aussagekräftige Einschätzung der Mortalität für das vorliegende Anwendungsgebiet möglich.

Das Verzerrungspotenzial wird als niedrig eingeschätzt. Es ist nicht davon auszugehen, dass die Imbalancen in der Anfallshäufigkeit zu Studienbeginn einen Einfluss auf die Todesfälle während der Studie haben.

4.5 Morbidität

Für die Endpunktkategorie „Morbidität“ legte der pU mit Modul 4 Daten zu epileptischen Anfällen in unterschiedlichen Operationalisierungen (Häufigkeit schwerer motorischer, sekundärer, tertiärer und jeglicher Anfälle, Anzahl anfallsfreier Tage, längstes anfallsfreies Intervall) und Auswertungen mittels CGI (CGI-I, CGI-CSID, CGICA, CGI-C–Zielverhalten), CSHQ und ADAMS vor. Die detaillierte Bewertung der Endpunkte ist im Kapitel 2.3 abgebildet. Zusammenfassend wurden für die Nutzenbewertung die Endpunkte zur prozentualen Veränderung und Reduktion (Responderanalysen $\geq 25\%$, $\geq 50\%$ und $\geq 75\%$) der Häufigkeit und Anzahl anfallsfreier Tage jeglicher Anfälle sowie der CGI-I herangezogen. Der primäre Endpunkt (prozentuale Veränderung der Häufigkeit schwerer motorischer Anfälle) wurde ergänzend im Kapitel 3.3 dargestellt.

Epileptische Anfälle

Es zeigten sich keine statistisch signifikanten Effekte in sämtlichen als relevant eingeschätzten Analysen zur Veränderung der Häufigkeit jeglicher Anfälle. Bezogen auf die prozentuale Veränderung der Anfallshäufigkeit war eine Reduktion im Ganaxolon-Arm gegenüber dem Placebo-Arm zwar ausgeprägter als in den Responderanalysen bzw. den anfallsfreien Tagen, jedoch war das Ergebnis ebenfalls nicht statistisch signifikant.

Zu Baseline zeigten sich im Ganaxolon-Arm gegenüber dem Placebo-Arm u. a. Unterschiede in den Anfallshäufigkeiten jeglicher Anfälle (Median: 116 vs. 60), der Anzahl anfallsfreier Tage jeglicher Anfälle (Median: 0 vs. 16 Tage) und schwerer motorischer Anfälle (12 vs. 24 Tage) sowie hinsichtlich des Anteils von nicht gehfähigen Personen (90 % vs. 75 %). Inwieweit diese Unterschiede die Effektstärken und das Signifikanzniveau beeinflussen, lässt sich nicht abschließend beantworten.

Aufgrund fehlender spezifischer Angaben zu den Modalitäten der Anfallserhebung mittels Tagebuches, zur Klassifikationsprüfung durch ESC und SIF/DRF sowie zu einer möglichen Schulung ist die korrekte Klassifikation der Anfälle als schwer motorisch, sekundär und tertiär nicht gänzlich nachvollziehbar. Die darüber hinaus bestehenden Imblancen zu Baseline in der Anfallshäufigkeit führen insgesamt zu einem unklaren Verzerrungspotential der Ergebnisse. (Tabelle 7).

Globaler Betreuerereindruck (CGI-I)

Der globale klinische Eindruck der Eltern/Betreuenden (CGI-I) zeigte in der Responderanalyse zu jeglicher Verbesserung zu Woche 17 keine statistisch signifikanten Unterschiede zwischen Ganaxolon und Placebo.

Das Verzerrungspotential des CGI-I wird aufgrund des hohen Anteils fehlender Werte insbesondere im Ganaxolon-Arm zu Woche 17 und den differentiellen Unterschieden zwischen den Behandlungsarmen (14 % bzw. 8 %) sowie der Nichtberücksichtigung der fehlenden Werte als Non-Responder als hoch erachtet. Das Vorgehen stellt eine Verletzung des ITT-Prinzips dar und kann zu einer Überschätzung des Behandlungseffektes führen. Teilweise (im Placebo-Arm vollständig) lassen sich die fehlende Werte zwar über Studienabbrechende erklären, jedoch liegt im Ganaxolon-Arm darüber hinaus ein Anteil (10 %) an aus unbekannten Gründen fehlender Werte vor. Die durchgeführten Sensitivitätsanalysen zur Imputation fehlender Werte weisen jedoch auf keine abweichenden Ergebnisse hin.

4.6 Lebensqualität

Für die Endpunktkategorie Lebensqualität legte der pU Daten für das patientenberichtete Instrument QI-Disability vor. Die detaillierte Bewertung der Endpunkte ist im Kapitel 2.3 abgebildet. Für die Nutzenbewertung wurde die Responderanalyse mit einer Verbesserung um 15 % des QI-Disability herangezogen.

In den Responderanalysen zur Verbesserung zum Endpunkt QI-Disability zeigte sich kein statistisch signifikanter Unterschied im Gesamtscore zwischen Ganaxolon und Placebo.

Das Verzerrungspotential des QI-Disability wird analog zum CGI-I, aufgrund des hohen Anteils fehlender Werte zu Woche 17 im Gesamtscore (jeweils ca. 16 %) und der damit einhergehenden Verletzung des ITT-Prinzips als hoch erachtet. Die Sensitivitätsanalysen weisen jedoch ebenfalls auf keine maßgebliche Veränderung der Ergebnisse nach Imputation fehlender Werte hin.

4.7 Sicherheit

Die Auswertungen zur Sicherheit von Ganaxolon beziehen sich auf UE, die ab Verabreichung der Studienmedikation bis 2 Wochen nach Abschluss der Taper-Phase, während der 17-wöchigen DBP aufgetreten sind. Die Behandlungsdauer war mit jeweils 16,9 Wochen im Median zwischen den Gruppen vergleichbar.

Der pU verwendet eine studieneigene Schweregradeinteilung für die unerwünschten Ereignisse. Eine Begründung hierzu führt der pU nicht an, jedoch ist die Einteilung an die Standardklassifikation CTCAE angelehnt und wird als adäquat eingestuft. Der pU reicht post hoc eine Analyse unter Nichtberücksichtigung von epileptischen Anfällen als Progressionsereignis ein, welche für die vorliegende Nutzenbewertung berücksichtigt wurde. Eine Definition für die spezifischen Anfallsereignisse wird nicht vorgelegt, zudem ist nicht auszuschließen, dass weitere Symptome der Grunderkrankung in die UE-Erfassung eingehen.

Bezogen auf die UE unter Ausschluss von epileptischen Anfällen zeigen sich auf aggregierter Ebene keine statistisch signifikanten Unterschiede im Auftreten von schweren UE, SUE und Therapieabbrüchen aufgrund von UE zwischen Ganaxolon und Placebo. Es traten häufiger SUE als schwere UE auf, was auf Basis der Operationalisierung nicht plausibel erscheint. Die Inzidenz der schweren UE lag auf Ebene der Preferred Terms (PT) und Systemorganklassen (SOC) bei < 5 %. SUE traten auf Ebene des SOC „Infektionen und parasitäre Erkrankungen“ bei 3 (6,0 %) Personen im Ganaxolon-Arm und 2 (3,9 %) Personen im Placebo-Arm auf und waren somit etwas häufiger. Angaben zur statistischen Signifikanz (p-Wert) und zu Effektschätzern (RR, 95%-KI) wurden nicht berichtet. Nach Angaben des pU sollten keine SUE auf Ebene der SOC und PT aufgetreten sein. Eine Rationale für die Diskrepanz zwischen den Studienunterlagen liegt nicht vor.

Häufig berichtete UE jeglichen Schweregrades ($\geq 10\%$) beziehen sich auf die SOC „Erkrankungen des Nervensystems“, „Infektionen und parasitäre Erkrankungen“, „Erkrankungen des Gastrointestinaltrakts“, „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“, „Erkrankungen der Atemwege, des Brustraums und Mediastinums“, „Psychiatrische Erkrankungen“, „Stoffwechsel- und Ernährungsstörungen“ und „Erkrankungen der Haut und des Unterhautgewebes“. Statistisch signifikante Unterschiede in den UE jeglichen Schweregrades wurden zuungunsten von Ganaxolon für die zum SOC „Erkrankungen des Nervensystems“ gehörende PT „Somnolenz“ (RR = 2,30; 95%-KI [1,10; 4,79]) beobachtet. Es ist unklar, warum keine Effektschätzer auf Ebene der SOC berichtet werden. In Gesamtschau der Ergebnisse ist nicht davon auszugehen, dass sich auf Ebene der SOC ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen ergibt.

Das Verzerrungspotential der Ergebnisse zu den UE wird als unklar bewertet.

5 Anforderung an eine qualitätsgesicherte Anwendung

Die Angaben der Fachinformation sind zu berücksichtigen.

6 Zusammenfassung der Nutzenbewertung

Ganaxolon ist zugelassen für Zusatzbehandlung von epileptischen Anfällen im Zusammenhang mit einer CDKL5-Mangelerkrankung bei Patienten im Alter von 2 bis 17 Jahren. ZTALMY kann bei Patientinnen und Patienten ab 18 Jahren weiter angewendet werden. Die Nutzenbewertung von Ganaxolon basiert auf der zulassungsbegründenden Studie Marigold (1042-CDD-3001). Es handelt sich bei der Studie Marigold um eine multizentrische, randomisierte, doppelblinde, placebokontrollierte Phase-III-Studie zur Untersuchung der Sicherheit und Wirksamkeit von Ganaxolon gegenüber Placebo bei Patientinnen und Patienten ab 2 Jahren mit CDD.

Die Ergebnisse der in der Nutzenbewertung berücksichtigten Endpunkte der Studie werden in der folgenden Tabelle unter Berücksichtigung des Verzerrungspotentials zusammengefasst. Es wird dargestellt, ob unter Ganaxolon 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.

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

Tabelle 21: Zusammenfassende Darstellung der Ergebnisse der Studie Marigold

Studie Marigold Darstellung der Ergebnisse	Ganaxolon N = 50			Placebo N = 51			Ganaxolon vs. Placebo	Effekt
Mortalität								
Todesfälle	N	Personen mit Ereignis, n (%)		N	Personen mit Ereignis, n (%)		RR [95%- KI]; p-Wert	
Todesfälle	50	0		51	0		-	↔
Morbidität								
Epileptische Anfälle (jegliche Anfälle)	N ¹⁾	Baseline- Phase ²⁾ Median [95%-KI] ³⁾	Prozentuale Veränderung ggü. Baseline ⁴⁾ n (%) Median [95%-KI] ³⁾	N	Baseline- Phase ²⁾ Median [95%-KI] ³⁾	Prozentuale Veränderung ggü. Baseline ⁴⁾ n (%) Median [95%-KI] ³⁾	Hodges- Lehmann- Schätzer [95%-KI] ³⁾ ; p-Wert ⁵⁾	
Veränderung der Häufigkeit ⁶⁾ jeglicher Anfälle ⁷⁾	49	116,0 [75,3; 173,3]	-19,1 [-36,8; - 3,28]	51	60,0 [46,0; 120,0]	-8,91 [-20,5; 9,60]	-17,4 [- 36,4; 0,33]; 0,051	↔
Veränderung der anfallsfreien Tage ⁸⁾ jeglicher Anfälle ⁷⁾	49	0,00 [0; 9,52]	0,07 [0; 2,51]	51	16,2 [9,52; 23,8]	-6,90 [-16,5; 15,3]	0,00 [-4,32; 2,56]; 0,78	↔
	N ¹⁾	Personen mit Ereignis, n (%)		N	Personen mit Ereignis, n (%)		RR [95%- KI] ⁹⁾ ; p- Wert ¹⁰⁾	
Reduktion der Häufigkeit ⁶⁾ jeglicher Anfälle ⁷⁾ um ≥ 25 %	49	22 (44,9)		51	15 (29,4)		1,63 [0,98; 2,72]; 0,07	↔

Studie Marigold Darstellung der Ergebnisse	Ganaxolon N = 50		Placebo N = 51		Ganaxolon vs. Placebo	Effekt
Reduktion der Häufigkeit ⁶⁾ jeglicher Anfälle ⁷⁾ um ≥ 50 %	49	10 (20,4)	51	4 (7,8)	2,60 [0,87; 7,75]; 0,09	↔
Reduktion der Häufigkeit ⁶⁾ jeglicher Anfälle ⁷⁾ um ≥ 75 %	49	5 (10,2)	51	1 (2,0)	5,20 [0,63; 43,0]; 0,11	↔
CGI-I (Eltern/ Betreuende)	<i>N¹¹⁾</i>	<i>Personen mit Ereignis, n (%)</i>	<i>N¹¹⁾</i>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%- KI]⁹⁾; p- Wert¹⁰⁾</i>	
Anteil der Personen mit jeglicher Verbesserung ¹²⁾	43	27 (62,8)	47	21 (44,7)	1,41 [0,95; 2,08]; 0,10	↔
Lebensqualität						
QI-Disability¹³⁾	<i>N¹⁴⁾</i>	<i>Personen mit Ereignis, n (%)</i>	<i>N¹⁴⁾</i>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%- KI]⁹⁾; p- Wert¹⁰⁾</i>	
Gesamtscore	42	4 (9,5)	43	3 (7,0)	1,32 [0,35; 4,99]; 0,71	↔
Sicherheit¹⁵⁾						
Unerwünschte Ereignisse	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%- KI]¹⁶⁾; p- Wert¹⁷⁾</i>	
Schwere UE ¹⁸⁾	50	1 (2,0)	51	1 (2,0)	1,02 [0,07; 15,9]; 1,00	↔
SUE	50	6 (12,0)	51	4 (7,8)	1,53 [0,46; 5,10]; 0,53	↔
UE, das zum Abbruch der Studienmedikation führte	50	1 (2,0)	51	2 (3,9)	0,51 [0,05; 5,45]; 1,00	↔

¹⁾ Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden. Fehlende Werte gehen nicht in die Analyse ein. Dies betrifft 1 Person (2,0 %) im Ganaxolon-Arm.

²⁾ Die Baseline-Phase umfasst die 6 Wochen vor Beginn der Behandlung mit der Studienmedikation (frühestens 42 Tage und spätestens 48 Tage nach Visite 1).

³⁾ Verteilungsfreies 95%-KI des Medians.

⁴⁾ Die Berechnung der prozentualen Änderung gegenüber Baseline erfolgte nur für Personen, deren Baselinewert nicht null war.

⁵⁾ p-Wert berechnet mit dem Wilcoxon-Rangsummen-Test.

⁶⁾ Angaben basieren auf der Summe einzelner Anfälle, zählbarer Anfälle und Cluster mit nicht zählbaren Anfällen (jedes Cluster mit nicht zählbaren Anfällen zählt als 1 Anfall). Die Häufigkeit der Anfälle ist normiert auf 28 Tage.

⁷⁾ Jegliche Anfälle umfassen sämtliche aufgetretene Anfallsarten.

- ⁸⁾ Der Anteil anfallsfreier Tage entspricht der Anzahl der Tage ohne Anfälle geteilt durch die Gesamtanzahl der Tage mit Anfallsdaten in den jeweiligen Studienphasen (DBP, Baseline), multipliziert mit 100.
- ⁹⁾ KIs berechnet mit der Wald-Methode.
- ¹⁰⁾ p-Wert berechnet mit dem exakten Fisher-Test.
- ¹¹⁾ Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden. Fehlende Werte gehen nicht in die Analyse ein. Der Anteil fehlender Werte beträgt 14,0 % im Ganaxolon-Arm und 7,8 % im Placebo-Arm. Es fand keine Erhebung zu Baseline statt.
- ¹²⁾ Jegliche Verbesserung war definiert als die Summe aus „geringfügig verbessert“, „stark verbessert“ und „sehr stark verbessert“.
- ¹³⁾ Die Skalenspannweite beträgt 0–100. Höhere Werte zeigen eine bessere Lebensqualität an.
- ¹⁴⁾ Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden. Fehlende Werte gehen nicht in die Analyse ein. Der Anteil fehlender Werte im Gesamtscore beträgt 16,0 % im Placebo-Arm und 15,7 % im Placebo-Arm. Zu Baseline betrug der Anteil fehlender Werte in der Analyse (basierend auf dem Gesamtscore) 4,0 % bzw. 2,0 %.
- ¹⁵⁾ Weitere Ergebnisse zur Sicherheit (UEs nach Organsystemen und Einzelereignisse) sind im Ergebniskapitel 3.5 Sicherheit dargestellt.
- ¹⁶⁾ Konfidenzintervalle post hoc berechnet mit der Wald-Methode.
- ¹⁷⁾ Effektschätzer post hoc berechnet; p-Wert post hoc berechnet mit dem exakten Fisher-Test.
- ¹⁸⁾ Es wurden studieneigene Kriterien zur Schweregradeinteilung verwendet:
- Leicht: UE, die leichte oder vorübergehende Beschwerden auslösen und eine minimale therapeutische Intervention oder Behandlung erfordern. Die alltäglichen Aktivitäten werden nicht beeinträchtigt.
- Moderat: UE, die tägliche Aktivitäten einschränken oder beeinträchtigen und durch spezifische therapeutische Interventionen gelindert werden können. Sie stellen jedoch kein signifikantes oder permanentes Risiko für die Patientinnen und Patienten dar.
- Schwer: UE, die zu erheblichen Einschränkungen des Gesundheitszustandes und Aktivitäten des täglichen Lebens führen und eine intensive therapeutische Intervention erfordern.

Abkürzungen: CGI-I: Clinical Global Impression of Improvement; DBP: doppelblinde, placebokontrollierte Studienphase; KI: Konfidenzintervall; OLE: offene einarmige Studienphase (Open-Label Extension); QI-Disability: Quality of Life Inventory-Disability; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Referenzen

1. **Amin S, Monaghan M, Aledo-Serrano A, Bahi-Buisson N, Chin RF, Clarke AJ, et al.** International Consensus Recommendations for the Assessment and Management of Individuals With CDKL5 Deficiency Disorder. *Front Neurol* 2022;13:874695. <https://dx.doi.org/10.3389/fneur.2022.874695>.
2. **Benke TA, Demarest S, Angione K, Downs J, Leonard H, Saldaris J, et al.** CDKL5 Deficiency Disorder. In: Adam MP, Bick S, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A (Eds). *GeneReviews*. Seattle (USA): 1993.
3. **Downs J.** Development, validation and use of the Quality of Life Inventory-Disability (QID Disability). ASID 2019 Conference; 6.-8.11.2019; Adelaide (AUS). https://asid.asn.au/wp-content/uploads/2022/03/ASID2019_QOL_Jenny-Downs.pdf.
4. **Epstein A, Williams K, Reddihough D, Murphy N, Leonard H, Whitehouse A, et al.** Content validation of the Quality of Life Inventory-Disability. *Child Care Health Dev* 2019;45(5):654-659. <https://dx.doi.org/10.1111/cch.12691>.
5. **Esbensen AJ, Rojahn J, Aman MG, Ruedrich S.** Anxiety, Depression and Mood Scale (ADAMS) [online]. Mapi Research Trust; 2003. [Zugriff: 20.07.2026]. URL: <https://eprovide.mapi-trust.org/instruments/anxiety-depression-and-mood-scale>.
6. **Esbensen AJ, Rojahn J, Aman MG, Ruedrich S.** Reliability and validity of an assessment instrument for anxiety, depression, and mood among individuals with mental retardation. *J Autism Dev Disord* 2003;33(6):617-629. <https://dx.doi.org/10.1023/b:jadd.0000005999.27178.55>.
7. **European Medicines Agency (EMA).** Ztalmy: European public assessment report EMEA/H/C/005825/0000 [online]. 25.05.2023. Amsterdam (NED): EMA. [Zugriff: 13.07.2026]. URL: https://www.ema.europa.eu/en/documents/assessment-report/ztalmy-epar-public-assessment-report_en.pdf.
8. **Hermans H, Jelluma N, van der Pas FH, Evenhuis HM.** Feasibility, reliability and validity of the Dutch translation of the Anxiety, Depression And Mood Scale in older adults with intellectual disabilities. *Res Dev Disabil* 2012;33(2):315-323. <https://dx.doi.org/10.1016/j.ridd.2011.09.018>.
9. **Immedica Pharma.** ZTALMY 50 mg/ml Suspension zum Einnehmen [online]. 15.01.2026. Frankfurt/Main. [Zugriff: 13.07.2026]. URL: <https://www.fachinfo.de/fi/detail/025821/ztalmy-50-mg-ml-suspension-zum-einnehmen>.
10. **Immedica Pharma Germany.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Ganaxolon (Ztalmy), Zusatzbehandlung von CDD-induzierten epileptischen Anfällen bei Patienten ab 2 Jahren; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 26.06.2026.
11. **Jacoby P, Epstein A, Kim R, Murphy N, Leonard H, Williams K, et al.** Reliability of the Quality of Life Inventory-Disability Measure in Children with Intellectual Disability. *J Dev Behav Pediatr* 2020;41(7):534-539. <https://dx.doi.org/10.1097/DBP.0000000000000815>.
12. **Leonard H, Junaid M, Wong K, Demarest S, Downs J.** Exploring quality of life in individuals with a severe developmental and epileptic encephalopathy, CDKL5 Deficiency Disorder. *Epilepsy Res* 2021;169:106521. <https://dx.doi.org/10.1016/j.epilepsyres.2020.106521>.

13. **Leonard H, Szepe E, Junaid M, Wong K, Saldaris J, Downs J.** Exploring sleep challenges in CDKL5 Deficiency Disorder. *J Neurodev Disord* 2026;18(1).
<https://dx.doi.org/10.1186/s11689-025-09666-w>.
14. **Marinus Pharmaceuticals.** A double-blind, randomized, placebo-controlled trial of adjunctive ganaxolone treatment in children and young adults with cyclin-dependent kinase-like 5 (CDKL5) deficiency disorder (CDD) followed by long-term open-label treatment; (1042-CDD-3001); clinical study report, vers. 2.0 [unveröffentlicht]. 28.05.2021.
15. **Marinus Pharmaceuticals.** A double-blind, randomized, placebo-controlled trial of adjunctive ganaxolone treatment in children and young adults with cyclin-dependent kinase-like 5 (CDKL5) deficiency disorder (CDD) followed by long-term open-label treatment; (1042-CDD-3001); protocol and protocol amendments [unveröffentlicht]. 26.05.2020.
16. **Marinus Pharmaceuticals.** A double-blind, randomized, placebo-controlled trial of adjunctive ganaxolone treatment in children and young adults with cyclin-dependent kinase-like 5 (CDKL5) deficiency disorder (CDD) followed by long-term open-label treatment; (1042-CDD-3001); statistical analysis plan [unveröffentlicht]. 18.05.2021.
17. **Markovich AN, Gendron MA, Corkum PV.** Validating the Children's Sleep Habits Questionnaire Against Polysomnography and Actigraphy in School-Aged Children. *Front Psychiatry* 2014;5:188. <https://dx.doi.org/10.3389/fpsy.2014.00188>.
18. **Mori Y, Downs J, Wong K, Anderson B, Epstein A, Leonard H.** Impacts of caring for a child with the CDKL5 disorder on parental wellbeing and family quality of life. *Orphanet J Rare Dis* 2017;12(1):16. <https://dx.doi.org/10.1186/s13023-016-0563-3>.
19. **Owens JA, Spirito A, McGuinn M.** The Children's Sleep Habits Questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep* 2000;23(8):1043-1051
20. **Pestana Knight EM, Olson HE.** CDKL5 Deficiency Disorder: Some Lessons Learned 20 Years After the First Description. *Am J Intellect Dev Disabil* 2024;129(2):101-109.
<https://dx.doi.org/10.1352/1944-7558-129.2.101>.
21. **Rojahn J, Rowe EW, Kasdan S, Moore L, van Ingen DJ.** Psychometric properties of the Aberrant Behavior Checklist, the Anxiety, Depression and Mood Scale, the Assessment of Dual Diagnosis and the Social Performance Survey Schedule in adults with intellectual disabilities. *Res Dev Disabil* 2011;32(6):2309-2320.
<https://dx.doi.org/10.1016/j.ridd.2011.07.035>.
22. **Saldaris JM, Jacoby P, Leonard H, Benke TA, Demarest S, Marsh ED, et al.** Psychometric properties of QI-Disability in CDKL5 Deficiency Disorder: Establishing readiness for clinical trials. *Epilepsy Behav* 2023;139:109069.
<https://dx.doi.org/10.1016/j.yebeh.2022.109069>.
23. **Schlarb AA, Schwerdtle B, Hautzinger M.** Validation and psychometric properties of the German version of the Children's Sleep Habits Questionnaire (CSHQ-DE). *Somnologie - Schlafforschung und Schlafmedizin* 2010;14(4):260-266.
<https://dx.doi.org/10.1007/s11818-010-0495-4>.