

Nutzenbewertung



Gemeinsamer
Bundesausschuss

**von Arzneimitteln mit neuen Wirkstoffen nach
§ 35a SGB V**

**Bewertung von Arzneimitteln für seltene Leiden
nach § 35a Absatz 1 Satz 11 i.V.m.**

5. Kapitel § 12 Nr. 1 Satz 2 VerfO

Wirkstoff: Burosumab

Datum der Veröffentlichung: 2. Januar 2020

Inhaltsverzeichnis

Tabellenverzeichnis.....	3
Abbildungsverzeichnis.....	4
Abkürzungsverzeichnis	5
Hintergrund.....	6
1 Fragestellung.....	7
2 Auswahl und Methodik der relevanten Studien.....	8
2.1 Studienbasis für die Nutzenbewertung	8
2.2 Aufbau und Design der eingeschlossenen Studie	9
2.3 Endpunkte.....	15
2.3.1 Patientenrelevanz der Endpunkte.....	16
2.3.2 Charakterisierung der patientenrelevanten und ergänzend dargestellten Endpunkte	17
2.4 Statistische Methoden	32
2.5 Verzerrungspotential auf Studien- und Endpunktebene	34
3 Ergebnisse der eingeschlossenen Studie.....	36
3.1 Studiencharakteristika und Studienmedikation	36
3.2 Mortalität	42
3.3 Morbidität	42
3.4 Lebensqualität	45
3.5 Sicherheit.....	45
3.6 Subgruppen	49
4 Diskussion der Methodik und Ergebnisse.....	50
4.1 Zulassungsstatus und Zulassungspopulation von Burosumab.....	50
4.2 Design und Methodik der Studie	50
4.3 Mortalität	52
4.4 Morbidität	52
4.5 Lebensqualität	54
4.6 Sicherheit.....	54
5 Anforderung an eine qualitätsgesicherte Anwendung	56
6 Zusammenfassung der Nutzenbewertung	57
Referenzen	59

Tabellenverzeichnis

Tabelle 1:	Übersicht über die Studienbasis	8
Tabelle 2:	Charakterisierung der Studie 301	9
Tabelle 3:	Charakterisierung der Interventionen der Studie 301	14
Tabelle 4:	Bewertung der Patientenrelevanz der Endpunkte der Studie 301	16
Tabelle 5:	Beschreibung der Endpunkte der Kategorie Mortalität	17
Tabelle 6:	Beschreibung der Endpunkte der Kategorie Morbidität	18
Tabelle 7:	Beschreibung der Endpunkte der Kategorie Lebensqualität	28
Tabelle 8:	Beschreibung der Endpunkte der Kategorie Sicherheit.....	29
Tabelle 9:	Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der aktiv kontrollierten Behandlungsperiode der Studie 301	32
Tabelle 10:	Analysepopulationen	33
Tabelle 11:	Verzerrungspotential in der Behandlungsperiode der Studie 301.....	34
Tabelle 12:	Verzerrungspotential in der Behandlungsperiode der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie 301	35
Tabelle 13:	Allgemeine Angaben zum Patientenfluss in Studie 301 mit Datenschnitt zu Woche 64.....	36
Tabelle 14:	Charakterisierung der FAS-Studienpopulation ¹⁾ der Studie 301 zu Baseline.....	37
Tabelle 15:	Exposition mit der Studienmedikation in Studie 301	39
Tabelle 16:	Begleitmedikationen zum Datenschnitt zu Woche 64 bei > 10 % der Patientinnen und Patienten in einer der Behandlungsgruppen der Studie 301	40
Tabelle 17:	Wesentliche Protokollverletzungen in Studie 301 bis zum Datenschnitt zu Woche 64	41
Tabelle 18:	Veränderungen in der Körpergröße zu Woche 64 in Studie 301, FAS-Population...	42
Tabelle 19:	Veränderung im 6MWT zu Woche 64 in Studie 301, FAS-Population ≥ 5 Jahre	43
Tabelle 20:	PROMIS-Scores zu Woche 64 in Studie 301, FAS-Population ≥ 5 Jahre.....	44
Tabelle 21:	Veränderung im FPS-R zu Woche 64 in Studie 301, FAS-Population ≥ 5 Jahre	45
Tabelle 22:	Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Behandlungsperiode bis Woche 64, Safety Set.....	46
Tabelle 23:	UE mit Inzidenz bei > 10 % der Patientinnen und Patienten in einer der Behandlungsgruppen während der Behandlungsperiode bis Woche 64 der Studie 301, Safety Set	46
Tabelle 24:	Dentale Ereignisse während der Behandlungsperiode in Studie 301, Safety Set	48
Tabelle 25:	UE mit CTCAE-Grad 3 während der Behandlungsperiode in Studie 301, Safety Set.....	48
Tabelle 26:	SUE während der Behandlungsperiode in Studie 301, Safety Set.....	49
Tabelle 27:	UE von besonderem Interesse während der Behandlungsperiode in Studie 301, Safety Set.....	49
Tabelle 28:	Zusammenfassende Darstellung der Ergebnisse der Studie UX023-CL301 für die Behandlungsperiode	57

Abbildungsverzeichnis

Abbildung 1: Patientenfluss in Studie 301 (Datenschnitt zu Woche 64 vorgelegt; Behandlungsphase von allen Testpersonen abgeschlossen) [21].....	13
Abbildung 2: Faces Pain Scale - Revised [29].....	25

Abkürzungsverzeichnis

6MWT	6-Minuten-Gehtest (Six-Minute Walk Test)
COA	Clinical Outcome Assessment
CTCAE	Common Terminology Criteria for Adverse Events
EMA	European Medicines Agency
EPAR	European Public Assessment Report
FAS	Full Analysis Set
FPS-R	Faces Pain Scale - Revised
G-BA	Gemeinsamer Bundesausschuss
GEE	Generalized Estimation Equation
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KI	Konfidenzintervall
PHEX	Phosphate regulating endopeptidase homolog X-linked
PROMIS	Patient-Reported Outcomes Measurement Information System
pU	pharmazeutischer Unternehmer
RGI-C	Radiographic Global Impression of Change
RSS	Rickets Severity Scale
SAP	Statistischer Analyseplan
SF-10	Short Form Health Survey-10 for Children
SGB	Sozialgesetzbuch
SUE	Schwerwiegende/s unerwünschte/s Ereignis/se
UE	Unerwünschte/s Ereignis/se
VerfO	Verfahrensordnung des G-BA
XLH	X-chromosomale Hypophosphatämie (X-linked hypophosphatemia)

Hintergrund

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

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

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

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

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Burosumab in seiner Sitzung am 18. Dezember 2019 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 26. September 2019 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pU in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 2. Januar 2020 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1 Fragestellung

Burosumab (CRYSVITA ®) 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 G-BA bestimmt bei Orphan Drugs, die einen Umsatz von 50 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. In die Bewertung werden gemäß Zulassungsstatus Patientinnen und Patienten in folgender Indikation eingeschlossen:

Behandlung von Kindern ab 1 Jahr und Jugendlichen in der Skelettwachstumsphase mit X-chromosomaler Hypophosphatämie (XLH) und röntgenologischem Nachweis einer Knochenerkrankung.

Burosumab durchlief für diese Indikation bereits 2018 eine Nutzenbewertung als Orphan Drug [9], welche im Beschluss vom 4. Oktober 2018 [10] mündete. Die Nutzenbewertung erfolgte auf Grundlage der zwei zulassungsbegründenden, offenen Phase-II-Studien UX023-CL201 (5–12 Jahre; Datenschnitt zu Woche 64) und UX023-CL205 (1–4 Jahre; Datenschnitt zu Woche 40). Im Stellungnahmeverfahren reichte der pU zudem Ergebnisse der randomisiert kontrollierten Studie UX023-CL301 mit Datenschnitt zu Woche 40 ein. Die finalen Ergebnisse dieser Studie waren der European Medicines Agency (EMA) im Juli 2019 vorzulegen. Da die finalen Ergebnisse auch für die Nutzenbewertung nach § 35a SGB V als relevant eingestuft wurden, wurde eine Befristung der Geltungsdauer des Beschlusses bis zum 1. Oktober 2019 als hinreichend erachtet.

2 Auswahl und Methodik der relevanten Studien

2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studie	Studie mit dem Dossier vom pU eingereicht	Studie vom pU als relevant für Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzenbewertung	Gründe, warum Studie nicht relevant für die Nutzenbewertung ist
Studien zum Wirkstoff				
UX023-CL301 ¹⁾	Ja	Ja	Ja	-
UX023-CL201 ²⁾	Ja	Nein	Nein	Diese offene, unkontrollierte Dosisfindungsstudie der Phase II mit Kindern zwischen 6 und 12 Jahren (N = 26 mit relevantem Dosierungsschema) bildet die Grundlage der Zulassung unter „Besonderen Bedingungen“, welche 2018 von der EMA gewährt wurde. Es liegt der Studienbericht zu Woche 160 vor. Da mit Studie UX023-CL301 für diese Population in allen Endpunktkategorien Evidenz mit einer Kontrollgruppe bis Woche 64 verfügbar ist, ist diese Studie nicht relevant zur Ableitung des Zusatznutzens.
UX023-CL205 ²⁾	Ja	Nein	Nein	Diese unkontrollierte Studie der Phase II mit Kindern zwischen 1 und 4 Jahren (N = 13) bildet die Grundlage der Zulassung unter „Besonderen Bedingungen“, welche 2018 von der EMA gewährt wurde. Es liegen der Studienbericht zu Woche 40 sowie statistische Outputs für Analysen bis Woche 64 vor. Da mit Studie UX023-CL301 für diese Population in allen Endpunktkategorien Evidenz mit einer Kontrollgruppe bis Woche 64 verfügbar ist, ist diese Studie nicht relevant zur Ableitung des Zusatznutzens.
KRN23-003	Ja	Nein	Nein	Diese unkontrollierte Studie der Phase III untersuchte u. a. die Selbstadministration von Burosumab bei Kindern zwischen 1 und 12 Jahren (N = 15). Es liegt der Studienbericht bis Woche 40 vor. Da mit Studie UX023-CL301 für diese Population in allen Endpunktkategorien Evidenz mit einer Kontrollgruppe bis Woche 64 verfügbar ist, ist diese Studie nicht relevant zur Ableitung des Zusatznutzens.
Andere Studien				
UX023-CL001	Ja	Nein	Nein	Diese nicht-interventionelle Querschnittsstudie hatte laut Studienbericht das Ziel, ein besseres Verständnis des XLH-Krankheitsverlaufs bei Kindern und Erwachsenen zu erreichen. Es können keine relevanten Erkenntnisse zum Zusatznutzen für Burosumab abgeleitet werden.

Studie	Studie mit dem Dossier vom pU eingereicht	Studie vom pU als relevant für Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzenbewertung	Gründe, warum Studie nicht relevant für die Nutzenbewertung ist
UX023-CL002	Ja	Nein	Nein	Diese nicht-interventionelle Studie sollte der pivotalen Zulassungsstudie UX023-CL201 als historischer Vergleich dienen. Da mit der direkt vergleichenden Studie UX023-CL301 höherwertige Evidenz verfügbar ist, ist diese Studie nicht relevant zur Ableitung des Zusatznutzens.

¹⁾ Wurde von der EMA 2018 als eine Bedingung der Zulassung unter „Besonderen Bedingungen“ genannt (s. [6]).

²⁾ Zulassungsrelevante Studie gemäß EPAR [6] für eine Zulassung unter „Besonderen Bedingungen“ (s. u.).

Abkürzungen: EMA: European Medicines Agency; EPAR: European Public Assessment Report; pU: pharmazeutischer Unternehmer; XLH: X-chromosomale Hypophosphatämie.

Burosumab wurde 2018 durch die EMA unter „Besonderen Bedingungen“ zugelassen. Als Bedingung der Zulassung wurde u. a. die Vorlage der finalen Ergebnisse aus den Studien UX023-CL201 und UX023-CL205 sowie von zum Zeitpunkt der Zulassung noch nicht verfügbaren Daten der Studie UX023-CL301 zum Nachweis der Wirksamkeit auferlegt. Es wurden keine spezifischen Auflagen zur Vorlage fehlender Sicherheitsdaten bestimmt [6].

Zur Nutzenbewertung für Burosumab wurden folgende Studien und Daten herangezogen:

- Herstellerdossier zu Burosumab [21],
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [6],
- Studienbericht [32] inkl. Studienprotokoll [31] und statistischem Analyseplan (SAP) [33] der Studie UX023-CL301,
- Publikation zum Wirkstoff: Imel et al. 2019 [13].

2.2 Aufbau und Design der eingeschlossenen Studie

Die Unterlagen zur Nutzenbewertung für Burosumab basieren auf Studie UX023-CL301, im Folgenden Studie 301 genannt. Die Studie und die Intervention werden in den Tabellen 2 und 3 charakterisiert.

Tabelle 2: Charakterisierung der Studie 301

Charakteristikum	Beschreibung
Design	<p>Die Studie UX023-CL301 ist eine multizentrische, randomisiert kontrollierte, offene Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit von Burosumab im Vergleich zur oralen Einnahme von Phosphat und aktivem Vitamin D bei Kindern mit XLH im Alter von 1 bis ≤ 12 Jahren. Primäres Zielkriterium der Studie war die Verbesserung der Rachitis basierend auf dem RGI-C-Gesamtwert von Baseline zu Woche 40.</p> <p>Die Randomisierung erfolgte stratifiziert nach den folgenden Faktoren:</p> <ul style="list-style-type: none"> • Rachitisschwere (RSS ≤ 2,5 vs. > 2,5) • Alter (< 5 vs. ≥ 5 Jahre) • Region (Japan vs. Rest der Welt).

Charakteristikum	Beschreibung
	<p>Die Studie besteht aus</p> <ul style="list-style-type: none"> • einer Screeningperiode (einschließlich einer einwöchigen Phosphat- und Vitamin-D-Wash-out-Periode), • einer 64-wöchigen kontrollierten Behandlungsperiode und • einer freiwilligen, bis zu 76-wöchigen Extensionsperiode (nur in Europa, Nordamerika und Australien), in der alle Patientinnen und Patienten Burosumab erhalten (s. Abbildung 1). <p>Die maximale Studiendauer beläuft sich auf 140 Wochen.</p> <p>Die Studie ist noch nicht abgeschlossen. Die Ergebnisse der Analyse zu Woche 64 mit Datenschnitt am 30.07.2018 liegen vor. Zum Zeitpunkt dieses Datenschnitts haben alle Studienteilnehmenden die Behandlungsperiode abgeschlossen.</p>
Population	<p>Wesentliche Einschlusskriterien</p> <ul style="list-style-type: none"> • Mädchen und Jungen im Alter von 1 bis 12 Jahren mit radiologischem Nachweis einer Rachitis, definiert als RSS-Gesamtscore von $\geq 2,0$ • PHEX-Mutation (oder Variante mit unklarer Bedeutung) bei der Studienperson oder einem direkt verwandten Familienmitglied mit entsprechender X-chromosomaler Vererbung • Biochemische Befunde, die mit XLH assoziiert sind: Serumphosphat $< 3,0$ mg/dl (0,97 mmol/l) • Serum-Kreatinin unterhalb der altersadjustierten oberen Normalbereichsgrenze • Serum 25(OH)D über dem unteren Normwert (≥ 16 ng/ml) zur Screening-Visite. (Falls 25(OH)D unter dem Normwert liegt, wird eine 25(OH)D-Substitution verschrieben. Unter der Annahme, dass die Testperson sonst alle anderen Einschlusskriterien erfüllt, war ein Re-Screening nach einem Minimum von 7 Tagen möglich) • Orale Einnahme von Phosphat und aktivem Vitamin D für ≥ 12 aufeinanderfolgende Monate (bei Kindern ≥ 3 Jahren) oder für ≥ 6 aufeinanderfolgende Monate (bei Kinder < 3 Jahren) bis 7 Tage vor der Randomisierung <p>Wesentliche Ausschlusskriterien</p> <ul style="list-style-type: none"> • Tanner-Stadium von 4 oder höher in einem der folgenden Merkmale: Genitalien, Brust oder Schamhaare • Perzentil der Körperhöhe > 50 basierend auf länderspezifischen Normen • Verwendung von Aluminiumhydroxid-Antazida, systemischen Kortikosteroiden, Acetazolamid und Thiaziden innerhalb von 7 Tagen vor Screening-Visite • Gegenwärtige oder frühere Verwendung von Leuprorelin, Triptorelin, Goserelin oder anderen Arzneimitteln, von denen bekannt ist, dass sie die Pubertät verzögern • Wachstumshormontherapie innerhalb von 12 Monaten vor Screening-Visite • Vorhandensein einer Nephrokalzinose im Nierenultraschall Grad 4 basierend auf folgender Skala: <ul style="list-style-type: none"> ○ 0 = normal ○ 1 = schwacher hyperechogener Rand um die Markpyramiden ○ 2 = intensiverer echogener Rand mit Echos, mit schwacher Füllung der Pyramide ○ 3 = gleichmäßig intensive Echos in der ganzen Pyramide ○ 4 = Steinbildung: solitärer Echo-Fokus an der Spitze der Pyramide • Geplante orthopädische Operation einschließlich Osteotomie, Implantation oder Entfernung von Klammern (Staples), 8-Platten oder anderen orthopädischen Materialien innerhalb der ersten 40 Wochen der Studie • Hypo- oder Hyperkalzämie, definiert als Serumkalziumspiegel außerhalb der altersadjustierten Normalwerte • Nachweis eines Hyperparathyreoidismus (PTH-Level 2,5-fach der oberen Normalbereichsgrenze) • Einnahme von PTH-Suppressiva innerhalb von 2 Monaten vor Screening-Visite

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> • Vorhandensein oder Vorgeschichte eines Zustandes, den das klinische Personal als ein hohes Risiko für eine schlechte Compliance oder einen möglichen Studienabbruch einstuft • Vorgeschichte einer rezidivierenden Infektion oder Veranlagung zur Infektion oder einer bekannten Immunschwäche • Verwendung eines monoklonalen Antikörpers innerhalb von 90 Tagen vor der Screening-Visite oder vorherige allergische oder anaphylaktische Reaktionen auf einen monoklonalen Antikörper • Gegenwärtige oder frühere Überempfindlichkeit gegen Burosumab-Hilfsstoffe, die nach Einschätzung des klinischen Personals mit einem erhöhten Risiko für unerwünschte Ereignisse einhergehen
Intervention und Zahl der Patientinnen und Patienten	<p>Es wurden 122 Testpersonen gescreent; davon wurden 61 randomisiert. Diese 61 Personen bilden die FAS-Population¹⁾.</p> <p>Behandlungsgruppen:</p> <ul style="list-style-type: none"> • Interventionsgruppe Burosumab: 29 Personen, davon 15 Personen ≥ 5 Jahre • Kontrollgruppe Orales Phosphat und aktives Vitamin D: 32 Personen, davon 20 Personen ≥ 5 Jahre <p>Vergleiche auch Abbildung 1.</p>
Ort und Zeitraum der Durchführung	<p>Ort der Durchführung 16 Studienzentren: USA (5), Kanada (3), Japan (3), Großbritannien (2), Schweden (1), Australien (1), Korea (1)</p> <p>Studiendauer</p> <ul style="list-style-type: none"> • Einschluss erste Patientin / erster Patient: 08.09.2016 • Einschluss letzte Patientin / letzter Patient: Keine Angabe • Aktueller Datenschnitt: 30.07.2018 (Datenschnitt zu Woche 64) Die a priori geplante Analyse zu Woche 64 fand statt, nachdem die letzte Testperson Studienwoche 64 durchlaufen hatte. • Geplantes Studienende: September 2018 / Juli 2019 (Ende der Extensionsperiode, je nach länderspezifischer Verfügbarkeit von Burosumab außerhalb von Studienprogrammen)
Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte gemäß Studienprotokoll	<p>Primärer Endpunkt Vergleich der Veränderungen im RGI-C-Gesamtwert von Baseline zu Woche 40 zwischen Interventions- und Kontrollarm</p> <p>Sekundäre Endpunkte</p> <ul style="list-style-type: none"> • Anteil der Personen mit RGI-C-Gesamtwert $\geq +2,0$ zu Woche 40 und Woche 64 • Veränderung der Rachitis basierend auf dem RGI-C-Gesamtwerts zu Woche 64 • Veränderung im RSS-Gesamtwert von Baseline zu Woche 40 und Woche 64 • Veränderung der Skelettanomalien der unteren Extremitäten basierend auf dem RGI-C-Bein-Score zu Woche 40 und 64 • Veränderung in der Körpergröße von Baseline zu Woche 24, 40 und 64 • Veränderung im altersnormierten z-Score der Körperhöhe/Liegegröße zu Woche 24, 40 und 64 • Veränderung in der Wachstumsgeschwindigkeit • Schmerz, Fatigue und körperliche Funktion: Veränderung in den PROMIS-Domänenwerten „Pediatric Pain Interference“, „Physical Function Mobility“ und „Fatigue“ (bei Personen ≥ 5 Jahre) zu Woche 24, 40 und 64 • Schmerzintensität: Veränderung im FPS-R (bei Personen ≥ 5 Jahre) zu Woche 24, 40 und 64 • Gehfähigkeit: Veränderung in der 6MWT-Gehstrecke und Anteil der erwarteten 6MWT-Gehstrecke (bei Personen ≥ 5 Jahre) zu Woche 24, 40 und 64

Charakteristikum	Beschreibung
	<p>Explorative Endpunkte</p> <ul style="list-style-type: none"> • Gesundheitsbezogene Lebensqualität mittels SF-10 (für Kinder ab 5 Jahren) • Dentale Ereignisse: Zahnkaries, verzögerter Durchbruch der Zähne, Zahnschmelzhypoplasie, Zahnabszess, Zahnextraktion (durch Abszess oder Karies), Wurzelbehandlung und/oder Gingivitis <p>Pharmakodynamische Endpunkte</p> <ul style="list-style-type: none"> • Phosphat, 1,25(OH)₂D im Serum • Phosphat im Urin, Renale Phosphatschwelle (TmP/GFR) und prozentuale tubuläre Phosphatrückresorption (TRP) • Knochenmarker <p>Pharmakokinetik Burosumab im Serum</p> <p>Sicherheit Inzidenz der UE und SUE</p>
<p>Subgruppenanalysen</p>	<p>Subgruppen gemäß SAP Für Wirksamkeitsendpunkte einschließlich Rachitisschwere, Wachstum und pharmakodynamische Endpunkte waren die folgenden Subgruppenanalysen a priori vorgesehen:</p> <ul style="list-style-type: none"> • RSS-Gesamtwert: ≤ 2,5 vs. > 2,5 • Alter: < 5 vs. ≥ 5 Jahre • Region: Japan vs. Rest der Welt • Nur Rachitisschwere und Wachstum: Geschlecht: weiblich vs. männlich <p>Zudem konnten je nach Angemessenheit²⁾ Subgruppenanalysen für PHEX-Mutationsstatus (positiv vs. nicht-positiv) und bei PROMIS, FPS-R und 6MWT nach RSS-Gesamtwert (≤ 2,5 vs. > 2,5) durchgeführt werden.</p> <p>Subgruppen für Modul 4 Der pU legt zudem für <u>alle</u> Wirksamkeits- und Sicherheitsendpunkte mit Modul 4 Subgruppenanalysen für die folgenden Merkmale vor:</p> <ul style="list-style-type: none"> • RSS-Gesamtwert: ≤ 2,5 vs. > 2,5 • Geschlecht: weiblich vs. männlich • Region: Japan vs. Rest der Welt • Alter: < 5 vs. ≥ 5 Jahre • PHEX-Mutationsstatus: positiv vs. nicht-positiv • Orthopädische Chirurgie während der Studie: ja vs. nein

¹⁾ Als FAS-Population waren alle Patientinnen und Patienten definiert, die mindestens eine Dosis der Studienmedikation erhielten. Da dies auf alle randomisierten Personen der Studie UX023-CL301 zutrifft, entspricht die FAS- der ITT-Population.

²⁾ Es sind keine Kriterien beschrieben, woran eine Angemessenheit bestimmt wurde.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; FAS: Full Analysis Set; FPS-R: Faces Pain Scale - Revised; GFR: glomerulären Filtrationsrate; ITT: Intention-to-Treat; PHEX: Phosphate regulating endopeptidase homolog X-linked; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; RGI-C: Radiographic Global Impression of Change; RSS: Rickets Severity Scale; SAP: Statistischer Analyseplan; SF-10: Short Form Health Survey-10 for Children; (S)UE: (Schwerwiegende/s) Unerwünschte/s Ereignis/se; TmP: maximale tubuläre Rückresorption von Phosphat; TRP: prozentuale tubuläre Phosphatrückresorption; XLH: X-chromosomale Hypophosphatämie.

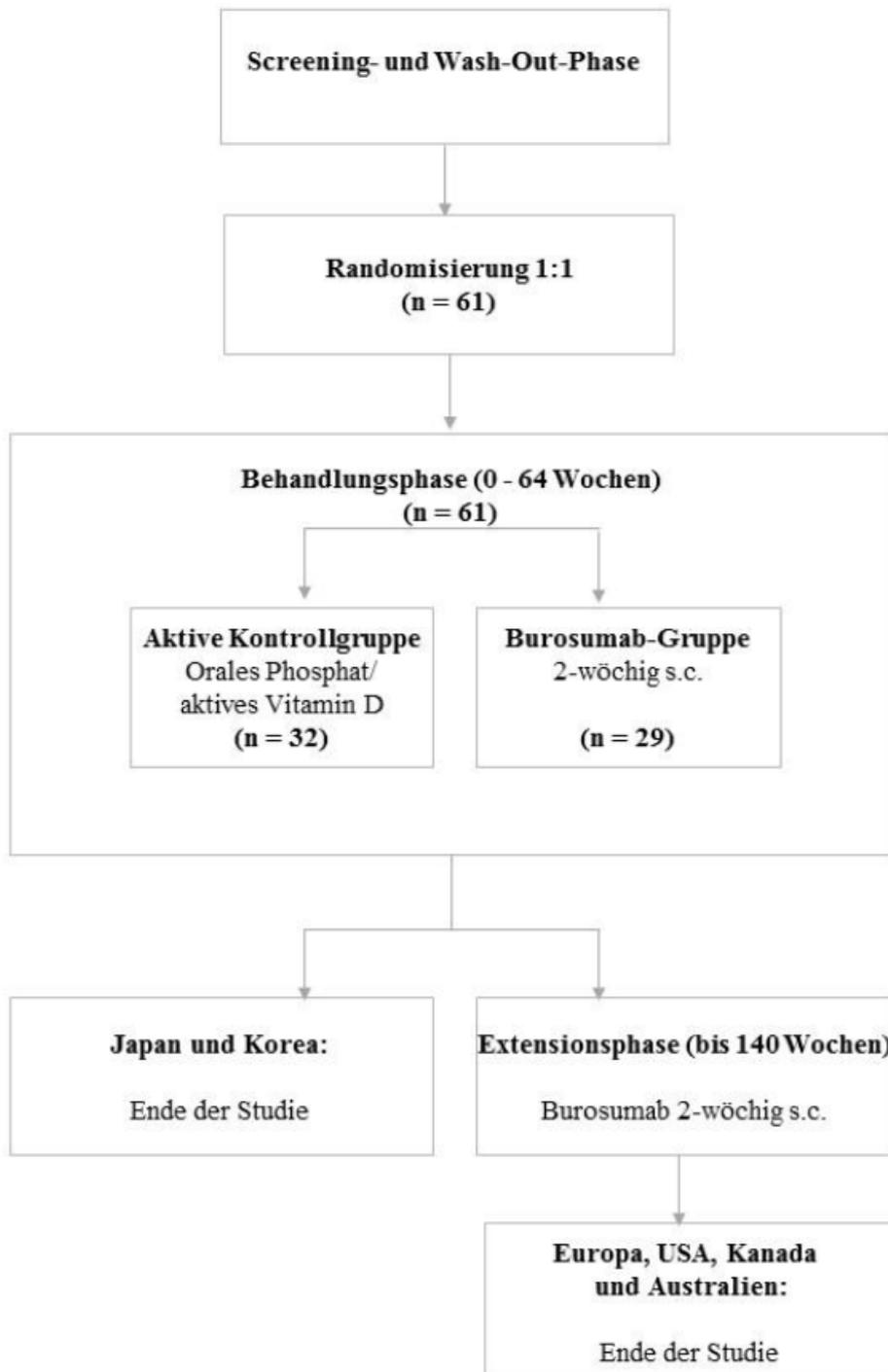


Abbildung 1: Patientenfluss in Studie 301 (Datenschnitt zu Woche 64 vorgelegt; Behandlungsphase von allen Testpersonen abgeschlossen) [21]

Tabelle 3: Charakterisierung der Interventionen der Studie 301

Intervention	Kontrolle ¹⁾
Burosumab <ul style="list-style-type: none"> • Dosierung: <ul style="list-style-type: none"> ○ 0,8 mg/kg Körpergewicht alle 2 Wochen ○ Dosiserhöhung auf 1,2 mg/kg Körpergewicht basierend auf Nüchtern-Serumphosphatwert möglich²⁾ • Applikation: Subkutane Injektion in Bauch, Oberarme, Oberschenkel oder Gesäß (Injektionsstelle wechselt bei jeder Injektion). Übersteigt die Dosis ein Volumen von 1,5 ml, sollte die Dosis an zwei Injektionsstellen verabreicht werden 	Orales Phosphat und aktives Vitamin D <ul style="list-style-type: none"> • Dosierung: individuell nach Ermessen des klinischen Studienpersonals auf der Grundlage US-amerikanischer und europäischer Leitlinien <ul style="list-style-type: none"> ○ US-Leitlinie: Calcitriol 20–30 ng/kg/Tag aufgeteilt in 2–3 Dosen und Phosphat in einer Elementarphosphordosis von 20–40 mg/kg/Tag aufgeteilt in 3–5 Dosen [3] ○ Europäische Leitlinie: Alfacalcidol 1–2 µg/Tag und Phosphat 45–70 mg/kg/Tag aufgeteilt in 3–4 Dosen [24]
Nicht erlaubte Begleitmedikation <ul style="list-style-type: none"> • Aluminiumhydroxid-Antazide, systemischen Kortikosteroide, Acetazolamid oder Thiazid • Wachstumshormone • Parathormon-Suppressiva • Jede Therapie mit Monoklonalen Antikörper (außer Studienmedikation) • Leuprorelin, Triptorelin, Goserelin oder andere Medikamente, von denen bekannt ist, dass sie die Pubertät verzögern • Patientinnen und Patienten der Burosumab-Gruppe durften während der gesamten Studie (kontrollierte Behandlungs- und Extensionsperiode) kein orales Phosphat und kein aktives Vitamin-D einnehmen; Patientinnen und Patienten der Kontrollgruppe durften während der Extensionsperiode kein orales Phosphat und kein aktives Vitamin-D einnehmen. Erlaubte Begleitmedikation Notwendige Begleitmedikation (abgesehen der nicht erlaubten Begleitmedikation) war gestattet, wobei keine signifikanten Änderungen der Medikation während der Studie vorgenommen werden sollten.	

¹⁾ Im Anschluss an die kontrollierte Behandlungsperiode (Woche 0–64) erhalten auch alle Personen in der Kontrollgruppe nach 7-tägiger Phosphat- und Vitamin-D-Wash-out-Periode in einer unkontrollierten Extensionsperiode bis zu 76 Wochen Burosumab.

²⁾ Anpassung der Burosumab-Dosis:

a) Kriterien zur Dosiserhöhung:

1) Zwei aufeinanderfolgende Serumphosphatmessungen liegen unter dem Normalbereich,
2) Serumphosphat ist um $\leq 0,5$ mg/dl gegenüber dem Ausgangswert gestiegen und
3) es wurde keine Dosis der Studienmedikation ausgelassen, die die Abnahme des Serumphosphats erklären würde.

b) Wenn das Serumphosphat über die obere Grenze des Normalbereichs steigt, werden die nachfolgenden Dosen zunächst ausgesetzt. Nach Ausschluss anderer Ursachen des erhöhten Serumphosphats wird die Burosumab-Behandlung mit der Hälfte der zuletzt erhaltenen Dosis unter Überwachung der Serumphosphatwerte fortgesetzt. Sofern die Kriterien unter a) erfüllt werden, wird die Behandlung mit der zuletzt erhaltenen vollen Dosis fortgeführt.

Protokolländerungen

Das Originalstudienprotokoll ist auf den 4. April 2016 datiert. Protokollamendment 1 trat am 3. November 2017 ca. 10 Monate nach Studienbeginn in Kraft. Relevante Änderungen umfassen folgende Aspekte:

- Einführung der einarmigen Extensionsphase (bis zu 76 Wochen) für Testpersonen in Europa, den USA, Kanada und Australien;
- Änderung der Nachbeobachtung der Sicherheitsendpunkte: Einschränkung der 12-wöchigen Nachbeobachtung auf lediglich diejenigen Personen, die nicht anderweitig mit Burosumab therapiert werden (z. B. durch Teilnahme an einer anderen Burosumab-Studie, s. Tabelle 5 und Tabelle 8);
- Erhöhung des Anteils von weiblichen Testpersonen auf 70 %;
- Präzisierung von Ein- und Ausschlusskriterien;

- Änderung der Versionen des Patient-Reported Outcomes Measurement Information System (PROMIS) von 1.0 auf 2.0;
- Definition des Endpunkts „Zahnärztliche Beurteilung“ als Karies, verzögerter Zahndurchbruch, Schmelzhypoplasie, Zahnabszess und Gingivitis.

Hinsichtlich der Änderung der Nachbeobachtung der Sicherheitsendpunkte legt der pU in Modul 4 des Herstellerdossiers dar, dass „[a]uch nach Amendment 1 [...] die Kinder, die anderweitig (nicht über die Studie UX023-CL301) Burosumab erhielten, dennoch nach Auflagen der Zulassungsbehörde beobachtet [wurden]. Dies betrifft Patienten, die nicht in die Extensionsphase aufgenommen werden konnten (Kinder aus Japan oder Korea)“ [21]. Unerwünschte Ereignisse (UE) während der Extensionsperiode in dieser Population werden daher außerhalb der Studie 301 erfasst.

Unklar ist, welcher Rationale die Änderung der Ein- und Ausschlusskriterien und die Erhöhung des Anteils der weiblichen Testpersonen in Amendment 1 folgte. Mit Datenschnitt im Februar 2018 zum Erreichen der Studienwoche 40 der zuletzt eingeschlossenen Testperson ergibt sich, dass die Patientenrekrutierung zum Zeitpunkt des Amendments 1 bereits abgeschlossen war.

Die Änderung der PROMIS-Version von 1.0 auf 2.0 wird in Amendment 1 damit begründet, die tatsächliche, bereits stattfindende Verwendung der aktualisierten Version von PROMIS in der Studie widerzuspiegeln. Auch laut Modul 4 des Herstellerdossiers [21] wurde nur Version 2.0 des PROMIS in der Studie verwendet. Damit ist davon auszugehen, dass in Amendment 1 nicht ausschließlich Änderungen für die Studiendurchführung prospektiv beschrieben wurden, sondern dass es sich ebenfalls um eine nachträgliche Dokumentation handelt. Diese Annahme würde durch die Darstellung der geänderten Patientenrekrutierung in Amendment 1 nach Abschluss der Patientenrekrutierung gestützt. Daraus resultierende Konsequenzen für die Studienbewertung werden unter 2.5 beschrieben.

Der pU hat zusätzlich ein Clinical Outcome Assessment (COA-)Dossier [30] eingereicht. Dieses beschreibt das Konzept und die Durchführung der Endpunkterhebungen im klinischen Entwicklungsprogramm zu Burosumab sowie eigene Untersuchungen zu den psychometrischen Eigenschaften der Instrumente im Erkrankungsbild. Konkreter Bezug wird auf Studie UX023-CL201 der Phase II sowie auf Studie 301 genommen. Es liegen die finale Version 5 vom 6. Juli 2017, Version 4 vom 24. April 2017 sowie Version 3 vom 23. November 2016 vor. In Studie 301 wurde die erste Testperson am 8. September 2016 eingeschlossen, sodass die Versionsänderungen während der laufenden Studie 301 und zwischen der Änderung des Studienprotokolls stattfanden. Der Stellenwert des COA-Dossiers gegenüber dem Studienprotokoll ist unklar. Zum Teil finden sich in den Dokumenten unterschiedliche Informationen zum Einsatz von Erhebungsinstrumenten (s. Tabelle 6). Die Änderungen zwischen den Versionen des COA-Dossiers können nur unzureichend und nicht für die Versionen 1 bis 3 beurteilt werden.

2.3 Endpunkte

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

Im folgenden Kapitel werden schrittweise die Eignung der vorgelegten Endpunkte hinsichtlich Patientenrelevanz, Operationalisierung, Validität, Verzerrungspotential und statistischer Auswertung beurteilt. Dazu wurden das Herstellerdossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt.

2.3.1 Patientenrelevanz der Endpunkte

Tabelle 4: Bewertung der Patientenrelevanz der Endpunkte der Studie 301

Endpunkt	Kategorie	Durch den pU als patientenrelevant bewertet	In der Nutzenbewertung als patientenrelevant bewertet
Mortalität ¹⁾	Mortalität	Ja	Ja
Rachitissymptomatik (RGI-C)	Morbidität	Ja	Nein ²⁾
Rachitissymptomatik (RSS)		Ja	Nein ²⁾
Serumphosphat		Ja	Nein ²⁾
Körpergröße (Steh-/Sitzhöhe bzw. Liegelänge)		Ja	Ja
Gehfähigkeit (6MWT)		Ja	Ja
PROMIS-Domänen: Pediatric Pain Interference Physical Function Mobility Fatigue		Ja	Ja
Schmerzintensität (FPS-R)		Ja	Ja
Dentale Ereignisse ¹⁾		Ja	Ja ³⁾
Short Form Health Survey-10 for Children (SF-10)	Gesundheitsbezogene Lebensqualität	Ja	Ja
Unerwünschte Ereignisse	Sicherheit	Ja	Ja ⁴⁾

¹⁾ Im Rahmen der Sicherheit erhoben.

²⁾ Die vom pU vorgelegten Nachweise sind nicht geeignet eine hinreichende Korrelation bzw. Validierung des Surrogatendpunktes auf patientenrelevante Endpunkte darzulegen.

³⁾ Ergänzend in der Kategorie Sicherheit dargestellt (s. Tabelle 6).

⁴⁾ Patientenrelevanz von Laborparametern unklar.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; FPS-R: Faces Pain Scale - Revised; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; RGI-C: Radiographic Global Impression of Change; RSS: Rickets Severity Scale.

Mit dem Dossier zu Burosumab wurden Unterlagen für Endpunkte vorgelegt, die in ihrer Patientenrelevanz in dieser Nutzenbewertung abweichend von der Einschätzung des pU bewertet werden. Zur Patientenrelevanz der Endpunkte Rachitissymptomatik mittels RGI-C (Radiographic Global Impression of Change) und RSS (Rickets Severity Scale) sowie des Endpunkts Serumphosphat wurden keine neuen Nachweise durch den pU vorgelegt, sodass die Einschätzung aus dem Nutzenbewertungsverfahren aus dem Jahr 2018 (Vorgangsnummer 2018-04-15-D-349) bestehen bleibt [8,9,10]. Die Endpunkte werden daher als nicht bewertungsrelevant zur Einschätzung des Zusatznutzens erachtet und in der Konsequenz nicht dargestellt.

2.3.2 Charakterisierung der patientenrelevanten und ergänzend dargestellten Endpunkte

Im Folgenden werden die als patientenrelevant eingeschätzten und ergänzend dargestellten Endpunkte (s. Tabelle 4) hinsichtlich ihrer Operationalisierung und Validität bewertet. Das Ergebnis der Betrachtung der Endpunkte findet sich aufgeteilt nach Endpunktkategorien in den Tabellen 5–8.

Tabelle 5: Beschreibung der Endpunkte der Kategorie Mortalität

Endpunktkategorie Mortalität	
Endpunkt	Mortalität
Operationalisierung	<p>Anzahl der Personen, die im Studienverlauf im Rahmen der Sicherheitserfassung als Todesfälle gemeldet wurden.</p> <p>Laut originalem Studienprotokoll wurden alle UE ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung bis zu 12 Wochen nach der letzten Gabe des Studienmedikaments aufgezeichnet. Im Studienprotokoll nach Amendment 1 findet sich die Einschränkung, dass nur Personen nachbeobachtet werden, die nicht an einer anderen Studie mit Burosumab teilnehmen und nicht mit kommerziell verfügbarem Burosumab oder mit Burosumab auf einem anderen Wege behandelt werden. UE würden laut pU jedoch außerhalb der Studie 301 weiterhin entweder nach einem anderen Studienprotokoll oder je nach Auflagen der Zulassung durch das Monitoring der Sicherheit gemeldet [21]. Im SAP findet sich zudem die Einschränkung, dass nur UE unter Behandlung (TEAE) ausgewertet werden, d. h. UE, die bei oder nach der ersten Dosis des Prüfprodukts auftraten und nicht bereits vor der ersten Dosisgabe bestanden.</p> <p>Erhebungszeitpunkte</p> <p>Jede Studienvisite. Weitere Hausvisiten waren im Burosumab-Arm je nach Nähe zum Studienzentrum und Patientenpräferenz während der Behandlungsperiode möglich (nicht in Japan und Südkorea). Laut Studienprotokoll wurden im Kontrollarm Sicherheitsereignisse außerhalb der Studienvisiten telefonisch erfragt. In der Extensionsperiode waren Hausvisiten für alle Studienteilnehmenden möglich (nicht in Japan und Südkorea).</p> <p>Für Studienteilnehmende, die die Behandlung nach Woche 64 nicht fortsetzten (z. B. durch Studienende in Japan und Korea) und nicht in einem anderen Rahmen weiterbeobachtet wurden, fand eine weitere Studienvisite nach 12 (\pm 1) Wochen nach der letzten Dosis statt.</p> <p>Auswertung</p> <p>Die Analysen beruhen auf dem Safety Set.</p>
Bewertung	<p>Operationalisierung</p> <p>Die Operationalisierung ist nachvollziehbar beschrieben.</p> <p>Validität</p> <p>In Studie 301 werden UE ab Unterzeichnung der Einverständniserklärung erfasst, jedoch erst ab der ersten Dosis der jeweiligen Studienmedikation ausgewertet. Es ist unklar, wieso UE erhoben, jedoch nicht ausgewertet werden. Der Endpunkt wird dennoch bis Woche 64 als valide bewertet.</p> <p>Durch die geplante selektive Nachbeobachtung im Rahmen der Studie 301 ist die Validität nach Woche 64 potentiell eingeschränkt. In der Extensionsperiode werden lediglich die Personen nachbeobachtet, die Burosumab im Rahmen der Studie 301 weitererhielten. Sofern die Patientinnen und Patienten aus der Studie nach Woche 64 ausschieden, was in Japan und Korea regelhaft der Fall war, konnte die Sicherheitserhebung auch außerhalb der Studie stattfinden. Für Studie 301 könnte dadurch eine mögliche Selektion in der Nachbeobachtung stattfinden.</p>

Endpunktkategorie Mortalität	
	<p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, <u>der Verlängerung des Überlebens</u>, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Gesamtüberleben ist daher patientenrelevant und wird in der Nutzenbewertung berücksichtigt.</p>

Abkürzungen: AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; TEAE: Treatment Emergent Adverse Events; UE: Unerwünschte/s Ereignis/se.

Tabelle 6: Beschreibung der Endpunkte der Kategorie Morbidität

Endpunktkategorie Morbidität	
Endpunkt	Körpergröße (Steh-/Sitzhöhe bzw. Liegelänge)
Operationalisierung	<p>Körpergröße wurde als Steh- und Sitzhöhe bei Personen ≥ 2 Jahre erfasst. Bei Kindern < 2 Jahren oder bei Kindern, die nicht in der Lage oder nicht bereit waren, für die Messung zu stehen, wurde die Liegelänge erhoben. Zu jedem Zeitpunkt wurde die Steh-/Sitzhöhe bzw. Liegelänge dreimal gemessen und der Mittelwert der Messungen für die Analysen verwendet. Die Messungen wurden von geschultem Personal durchgeführt und waren durch ein Manual [29] standardisiert. Das Manual wurde mit den Studienunterlagen eingereicht. Steh-/Sitzhöhe und Liegelänge wurden in cm dargestellt. Für Stehhöhe und Liegelänge wurden ebenfalls alters- und geschlechtsadjustierten z-Scores und Perzentilen dargestellt. Die z-Scores spiegeln die Anzahl der Standardabweichungen eines jeden Wertes von den nach Alter und Geschlecht standardisierten normalen Durchschnittswerten wider. Im SAP wurde festgelegt, dass als Vergleichsdaten der z-Scores eine repräsentative Stichprobe gesunder Kinder in den USA (= Wachstumscharts des Centers for Disease Control) verwendet werden. Diese sind nicht spezifisch für Kinder mit besonderen Behandlungsanforderungen. Die Perzentile basieren ebenfalls auf Daten aus den USA.</p> <p>Erhebungszeitpunkte Baseline, Studienwoche 24, 40, 64 (= Studienende Japan und Südkorea), 76, 88, 100, 112, 124, 140 (oder vorzeitiger Abbruch). Die Auswertung erfolgte auf Basis der FAS-Population für den Datenschnitt zu Woche 64. Um Aussagen zur klinischen Relevanz einer Veränderung ableiten zu können, legt der pU im Herstellerdossier post hoc Berechnungen eines Hedges' g vor.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nachvollziehbar beschrieben.</p> <p>Validität und Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Die anthropometrischen Parameter können insbesondere bei Kindern mit charakteristischen, krankheitsbedingten Wachstumsstörungen als patientenrelevante Morbiditätsparameter eingeschätzt werden. Dabei sind Angaben, welche für Alter und Geschlecht adjustiert wurden, gegenüber absoluten Werten bevorzugt heranzuziehen. Die ermittelten z-Scores und Perzentilen basieren auf einer Stichprobe gesunder Kinder aus den USA. Länderspezifische Referenzpopulationen wurden nicht berücksichtigt.</p>

Endpunktkategorie Morbidität	
Endpunkt	Gehfähigkeit (6MWT)
Operationalisierung	<p>Der 6MWT wurde in Studie 301 bei Kindern eingesetzt, die zu Baseline mindestens 5 Jahre alt waren.</p> <p>Laut Studienunterlagen wurde der Test grundsätzlich nach den Vorgaben der American Thoracic Society (ATS)-Statement [22] von geschultem klinischen Studienpersonal durchgeführt. Testpersonen wurden angewiesen, eine vorher abgemessene Strecke über 6 Minuten (wiederholt) zu laufen. Die zurückgelegte Strecke wurde nach Ablauf der 6 Minuten in Metern erfasst. Der Versuchsablauf war durch ein Manual standardisiert, welches mit den Studienunterlagen eingereicht wurde [29]. Im Manual wird eine modifizierte Durchführung des Tests im Vergleich zum ATS-Statement beschrieben, indem farbige Ringe während jeder Runde in eine Box abgeworfen werden konnten.</p> <p>Der pU machte im Rahmen des Herstellerdossiers Angaben zur Rationale dieser Modifikation: „Aufgrund des unsicheren Ganges der XLH-Patienten führt die Messung der Gehstrecke mittels Messrades zu einem potentiell erhöhten Sturzrisiko der Patienten. Folglich wurde auf ein Messrad verzichtet, um eine freie Bewegung der oberen Extremitäten und eine damit verbundene Gleichgewichts-anpassung zu ermöglichen. Zur Vorbeugung einer möglichen Motivationsverzerrung durch das fehlende Messrad, erhielten die Patienten die Möglichkeit an jeder Biegung Softringe zu sammeln sowie fallen zu lassen. Dementsprechend ist mit der Softring-Technik für die Kinder ein spielerischer Anreiz geschaffen, die 6 Minuten kontinuierlich zu gehen, um möglichst viele Ringe zu sammeln“ [21].</p> <p>Erhebungszeitpunkte</p> <p>Screeningvisite (zu Übungszwecken), Baseline, Studienwoche 24, 40, 64 (= Studienende Japan und Südkorea), 88, 112, 140 sowie innerhalb von 3 Monaten bei vorzeitiger Beendigung der Studie</p> <p>Die Auswertung erfolgte auf Basis der FAS-Population (≥ 5-Jährige) für den Datenschnitt Woche 64. Um Aussagen zur klinischen Relevanz einer Veränderung ableiten zu können, legt der pU im Herstellerdossier post hoc Berechnungen eines Hedges' g vor.</p>
Bewertung	<p>Operationalisierung</p> <p>Die Operationalisierung ist nachvollziehbar beschrieben. Es sei darauf hingewiesen, dass die Durchführung mit Messrad nicht Teil des ATS-Statements ist, sondern aus einer Publikation für Referenzwerte stammt (s. u.).</p> <p>Validität</p> <p>Der 6MWT wurde entwickelt zur Messung der funktionalen körperlichen Fähigkeiten bzw. körperlichen Fitness [22]. Es handelt sich um ein standardisiertes und etabliertes Testverfahren. Es wird grundsätzlich davon ausgegangen, dass bei Einhaltung der Vorgaben der ATS-Richtlinie, valide Testergebnisse bei XLH-Erkrankten erzielt werden können. Die Reproduzierbarkeit der Testergebnisse des 6MWT ist bei qualitätssichernden Maßnahmen hoch. Zu diesen Maßnahmen gehören z. B. Schulungen des Studienpersonals, standardisierte Formulierungen der Ermutigungen nach bestimmten Zeiteinheiten, Messung durch die gleiche Person, ein bis zwei Probetestläufe. Laut Manual fanden in Studie 301 diese Maßnahmen statt.</p> <p>In der ATS-Richtlinie werden zudem mehrere potentielle Einflussfaktoren bzw. Quellen für Variation der Testergebnisse des 6MWT genannt. Hierzu zählen zum einen Patientencharakteristika (Alter, Körpergröße und -gewicht, Geschlecht) als auch äußere Faktoren wie Motivation, Schulung des Studienpersonals und Erfahrung der Testpersonen mit dem 6MWT. Wird eine Veränderung der körperlichen Fähigkeiten durch den 6MWT ermittelt, sollten mögliche Einflussfaktoren konstant gehalten werden, um eine potentielle Veränderung der Gehstrecke auf den zu untersuchenden Faktor, in diesem Fall auf die Studienmedikation, zurückführen zu können.</p> <p>Der pU hat eigene Untersuchungen zur Konstruktvalidität des modifiziert durchgeführten 6MWT basierend auf den Daten der Phase-II-Dosisfindungsstudie UX023-CL201 durch-</p>

Endpunktkategorie Morbidität	
	<p>geführt. Ergebnisse des 6MWT unterschieden sich im Median signifikant zwischen Kindern</p> <ul style="list-style-type: none"> • mit einer Stehhöhe ≥ 2 Standardabweichungen unterhalb des Mittelwerts und < 2 Standardabweichungen unterhalb des Mittelwerts, • und mit einem BOT-2-Skalenwert „Kraft und Gewandtheit“ von ≤ 40 und > 40 [30]. <p>Kongruente Validität wurde ebenfalls anhand der Ergebnisse der Studie UX023-CL201 untersucht. Korrelationen des 6MWT mit zwei Untertests des BOT-2 waren mindestens moderat ausgeprägt. Die stärkste Korrelation wurde zum BOT-2-Untertest „Schnelligkeit und Geschicklichkeit“ zu Baseline (Pearsons $r = 0,858$), die schwächste Korrelation zum Untertest „Stärke“ zu Woche 40 (Pearsons $r = 0,574$) beobachtet [30].</p> <p>Gegenwärtig existieren Untersuchungen zur MID des 6MWT in unterschiedlichen Patientenpopulationen, welche zu heterogenen Ergebnissen gelangen [2,26,27]. Untersuchungen bei XLH-Erkrankten liegen nicht vor. Zu diesem Zeitpunkt kann keine abschließende Einschätzung vorgenommen werden, ob eine relative Veränderung (z. B. Prozent vom Baselinewert) und/oder absolute Veränderung der Gehstrecke (in Metern) eine klinisch relevante Veränderung adäquat abbilden. Es ist davon auszugehen, dass eine Schwelle für klinisch relevante Veränderungen der 6MWT-Gehstrecke erkrankungs- bzw. populationspezifisch abzuleiten ist, wobei die Art der Erkrankung, Erkrankungsschwere als auch individuelle Patientencharakteristika wie Alter, Geschlecht und Anthropometrie relevanten Einfluss auf die klinische Relevanzschwelle haben können.</p> <p>Der pU stellt einen Vergleich zu Normwerten dar, die aus einer Untersuchung von 3- bis 11-Jährigen gesunden Kindern/Jugendlichen [7] stammen. Diese Kinder absolvierten einen modifizierten 6MWT mithilfe eines Messgeräts (Messrad), welches den ständigen Fortschritt/Erfolg als Länge der zurückgelegten Strecke anzeigte. Es ist unklar, inwiefern diese Ausführung mit der modifizierten Durchführung des 6MWT in Studie 301 vergleichbar ist. Aufgrund des kontrollierten Studiendesigns liegen mögliche Verzerrungen in beiden Studienarm vor, sodass die Ergebnisse im Rahmen der Nutzenbewertung dargestellt werden.</p> <p>Patientenrelevanz</p> <p>Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p> <p>In der Nutzenbewertung werden daher die Ergebnisse des 6MWT berücksichtigt.</p>
Endpunkt	PROMIS-Domänen: „Paediatric Pain Interference“, „Physical Function Mobility“ und „Fatigue“
Operationalisierung	<p>Schmerzen, körperliche Funktion und Fatigue wurden in Studie 301 mittels PROMIS erhoben.</p> <p>PROMIS ist ein System, das aus domänenspezifischen Instrumenten zur selbst- und fremdberichteten Beurteilung des Wohlbefindens besteht. Der domänenspezifische Ansatz basiert auf der Idee, dass Gesundheitsmerkmale wie Schmerz und körperliche Funktion nicht auf eine bestimmte Krankheit beschränkt sind. Die Domänen sind als Itembanken organisiert, aus denen individuell Items ausgewählt werden können. Die pädiatrischen Itembanken (für Kinder im Alter von 5–17 Jahren) umfassen generische Instrumente zur Erfassung von Funktionen, Symptomen, Verhaltensweisen und Gefühlen. Es stehen sowohl Instrumente zur Selbsteinschätzung (für Kinder ab 8 Jahren) als auch zur Fremdbeurteilung (Elternversion bei Kindern zwischen 5 und 7 Jahren) zur Verfügung.</p> <p>Laut PROMIS-Manual besteht neben der individuellen, dynamischen Auswahl der Items aus den Itembanken (im Rahmen Computer-adaptiver Tests (CAT)) die Möglichkeit, statische Kurzformen zu verwenden. Bei Verwendung von statischen Kurzformen beantworten die Testpersonen alle vorgegebenen Items einer Fragebogenversion. Die</p>

Endpunktkategorie Morbidität

PROMIS-Entwicklungsgruppe schlägt für jede Itembank eine Kurzform vor. Eine eigene Itemauswahl zur Generierung einer benutzerdefinierten Kurzform ist ebenfalls möglich. Bei Verwendung des CAT-Systems werden die Items basierend auf den Antworten der Testpersonen computergestützt aus den Itembanken für jede Testperson individuell zusammengestellt. Dadurch können beim CAT die Items je nach Testperson variieren, die Scores sollen jedoch zwischen den Testpersonen vergleichbar sein [34].

In Studie 301 wurden Items aus den PROMIS-Itembanken der Domänen „Pain Interference“, „Physical Function Mobility“ und „Fatigue“ (Version 2.0) ausgewählt und eine statische Fragebogenversion zur Selbst- und Fremdeinschätzung generiert. Dabei entsprechen die Items nicht vollständig den Items der von der PROMIS-Entwicklungsgruppe vorgeschlagenen Kurzformen mit 8–10 Items.

- Pediatric Pain Interference Domain:
 - 4 Items zu Problemen beim Schlafen, Rennen, Gehen, Stehen aufgrund von Schmerzen
 - 5-Punkte-Likert-Skala (0–4) je Item: nie / fast nie / manchmal / oft / fast immer
- Physical Function Mobility Domain:
 - 10 Items zu körperlichen Aktivitäten, Alltagsbewegungen
 - 5-Punkte-Likert-Skala (0–4) je Item: ohne Schwierigkeiten / mit etwas Schwierigkeiten / mit einigen Schwierigkeiten / mit viel Schwierigkeiten / nicht in der Lage, es zu tun
- Fatigue Domain:
 - 8 Items zu Einschränkungen in sozialen, sportlichen und alltäglichen Aktivitäten aufgrund von Müdigkeit und Erschöpfung
 - 5-Punkte-Likert-Skala (0–4) je Item: nie / fast nie / manchmal / oft / fast immer

Der Bezugszeitraum ist jeweils die letzte Woche.

Der Einsatz des Instruments war in Studie 301 über ein Manual standardisiert, welches vom pU mit den Studienunterlagen vorgelegt wurde [29]. Der Fragebogen wurde bei Kindern ab 5 Jahren und zu Beginn jeder Studienvisite eingesetzt. Bei Kindern, die beim Screening zwischen 5 und 7 Jahre alt waren, erfolgte die Beurteilung zu allen Erhebungszeitpunkten durch die Eltern. Bei Kindern ≥ 8 Jahre zum Zeitpunkt des Screenings wurde die Version zur Selbsteinschätzung verwendet. Sofern ein Kind Probleme beim Ausfüllen des Fragebogens zur Selbsteinschätzung hatte, konnte mit Begründung durch das Studienpersonal auch die Elternversion eingesetzt werden. In dem Fall musste in allen folgenden Studienvisiten die gleiche Ausfüllpraxis eingehalten werden.

Bei Vergleich der eingereichten Fragenbögen zur Selbst- und zur Fremdeinschätzung unterschieden sich Items dieser Versionen nur in ihrer Wortwahl, nicht in der Bedeutung. Im Herstellermodul 4 findet sich jedoch die Aussage, dass sich in der Version zur Selbsteinschätzung zusätzlich eine Frage zur Schmerzintensität befindet. Auch laut eines COA-Dossier (Version 5.0) mit Stand vom 06.07.2017 enthielt die Version zur Selbsteinschätzung eine Frage zur Schmerzintensität. In Studienbericht und -protokoll finden sich wiederum keine Hinweise auf eine Erhebung der Schmerzintensität mittels PROMIS.

Scoring und Interpretation der Werte

Die Domänenscores wurden in der Studie als T-Werte dargestellt. Dazu verweist der pU auf das PROMIS Online Scoring System. Demnach ergeben sich aus der Summe aller Itemscores einer Domäne die rohen Domänen-Werte. Der T-Score skaliert den Domänenrohwert in einen standardisierten Score mit einem Mittelwert von 50 und einer Standardabweichung (SD) von 10. Grundsätzlich repräsentiert ein höherer PROMIS-T-Score mehr von dem zu messenden Konzept:

- Für negativ formulierte Konzepte wie bei „Pain Interference“ und „Fatigue“ ist ein T-Score von 60 eine SD schlechter und ein T-Score von 40 eine SD besser als der Durchschnitt. Das bedeutet, dass höhere Werte eine stärkere Beeinträchtigung durch Schmerzen bzw. eine stärkere Fatigue repräsentieren.

Endpunktkategorie Morbidität	
	<ul style="list-style-type: none"> Für positiv formulierte Konzepte wie bei „Physical Activity“ ist ein T-Score von 60 eine SD besser und ein T-Score von 40 eine SD schlechter als der Durchschnitt. Höhere Werte sind damit gleichbedeutend mit einer besseren körperlichen Funktion. <p>Die T-Werte der Proxy-Einschätzungen für die Kinder zwischen 5 und 7 Jahren wurden vom pU gemeinsam mit den T-Werten der Selbsteinschätzung der Kinder ab 8 Jahren ausgewertet.</p> <p>Der pU referenziert sich im Herstellerdossier auf eine MID von 2,4 bis 3,5 bei der Interpretation der Ergebnisse aus der Analyse von Gruppenunterschieden. Responderanalysen wurden nicht vorgelegt.</p> <p>Erhebungszeitpunkte Baseline, Studienwoche 24, 40, 64 (= Studienende Japan und Südkorea), 88, 112, 140 (oder vorzeitiger Abbruch). Die Auswertung erfolgte auf Basis der FAS-Population (≥ 5-Jährige) für den Datenschnitt zu Woche 64.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist weitestgehend nachvollziehbar.</p> <ul style="list-style-type: none"> Diskrepanzen bestehen bezüglich der PROMIS-Skala „Pain Intensity“, einer 10-Punkte-Rating-Skala zur Messung der Schmerzintensität. Laut Herstellermodul und COA-Dossier [30] enthielt die eingesetzte Fragebogenversion zur Selbsteinschätzung für Kinder ab 8 Jahre diese Skala. Allerdings sind in Protokoll und Studienbericht der Studie 301 keine Angaben zur Messung der Schmerzintensität anhand der PROMIS-Pain-Intensity zu finden. Vielmehr wird in diesen Studienunterlagen angegeben, dass die Schmerzintensität mittels des Instrumentes FPS-R erhoben wurde (s. u.). Falls sich die Fragebogenversionen zur Selbst- und zur Fremdeinschätzung in der Erfassung der Schmerzintensität unterschieden, ist unklar, wie diese gemeinsam ausgewertet werden sollten. Auf welche Referenzpopulation sich die T-Werte beziehen, konnte anhand der Angaben im Manual nicht eindeutig nachvollzogen werden. Die vorgelegten Analysen untersuchen Gruppenunterschiede. Eine validierte Irrelevanzschwelle zur Beurteilung dieses Unterschieds liegt nicht vor. Die vom pU vorgeschlagene MID ist nicht geeignet, um einen Gruppenunterschied hinsichtlich klinischer Relevanz zu beurteilen. Stattdessen wären Responderanalysen unter Verwendung einer MID möglich. In der zugehörigen Publikation wurde zudem eine MID im Bereich von 1,9 bis 2,3 abgeleitet [28]. <p>Validität Der pU begründet die Auswahl der Items anhand des Vergleichs der ausgewählten Items mit den in der Literatur beschriebenen Symptomen und Auswirkungen der XLH und den Ergebnissen qualitativer Interviews mit erkrankten Kindern und ihren Eltern. Basierend auf den Ergebnissen der Literaturrecherche und der Interviews wurde ein konzeptionelles Modell entwickelt und die in jeder der Skalen enthaltenen Items auf das konzeptionelle Modell abgebildet. Der pU führt aus, dass Schmerzen, Fatigue und Einschränkungen in den körperlichen Funktionen die Hauptsymptome und Auswirkungen der XLH bei Kindern und Jugendlichen sind und diese Aspekte mit den PROMIS-Items abgedeckt werden.</p> <p><u>Instrumente zur Selbsteinschätzung</u> Die Entwicklung der pädiatrischen PROMIS-Itembänke zur Selbsteinschätzung umfasste 6 Phasen: Identifikation vorhandener Items, Item-Klassifikation und -Auswahl, Item-Review und -Revision, Fokusgruppen-Input zur Domain-Abdeckung, kognitive Interviews mit Einzelpersonen und abschließende Überarbeitung. Das finale Itemset enthielt 293 Items in 6 Domänen (Physical Function, Emotional Distress, Social Role Relationships, Fatigue, Pain, Asthma) [18].</p>

Endpunktkategorie Morbidität

Der Pool aus 293 Items wurde in einer Population von gesunden und chronisch kranken Kindern und Jugendlichen im Alter von 8 bis 17 Jahren (N = 4.129) mittels Item Response Theory (IRT) getestet [17].

Zur Untersuchung wurden die 293 Items randomisiert auf mehrere Test-Fragebögen (mit zum Teil überlappenden Items) aufgeteilt, um die Anzahl der zu beantwortenden Fragen pro Testperson zu begrenzen. Die Items zu „Pain“, „Fatigue“ und „Physical Function“ wurden dabei auf 4 Test-Fragebögen (ca. 65 PROMIS-Items pro Fragebogen) verteilt und von jeweils mindestens 750 Kindern beantwortet [5,23,35]. Basierend auf diesem Datensatz wurden IRT-Analysen bezüglich Dimensionalität der Skalen, lokale Abhängigkeit der Items und Differential Item Functioning (DIF) durchgeführt. Items, die das Kriterium der Eindimensionalität erfüllten, wurden mittels IRT kalibriert. Die finalen Itembänke der PROMIS-Version 1.0 umfassten insgesamt 165 Items über 5 generische Gesundheitsdomänen und einer Asthma-Domäne. Davon waren 23 Items der Itembank „Physical Function Mobility“, 23 Items der Itembank „Fatigue-Tired“ und 13 Items der Itembank „Pain Interference“ zugeordnet [5,23,35].

Inzwischen liegt PROMIS in der Version 2.0 vor, bei der einige Itembänke um Items ergänzt wurden. Die aktuelle Itembank zu „Pediatric Pain Interference“ enthält 20 Items, die für „Fatigue“ 25 Items und die für „Physical Function Mobility“ 24 Items [11].

Die in der Studie verwendeten Items waren bereits in der PROMIS-Version 1.0 enthalten.

Untersuchungen zur Konstruktvalidität und Änderungssensitivität konnten für folgende Skalen der Kurzformen identifiziert werden: „Pediatric Pain Interference“ bestehend aus 8 Items, „Physical Function Mobility“ mit 8 Items und „Fatigue“ mit 10 Items [19]. Zu beachten ist, dass alle 4 in Studie 301 eingesetzten Pain-Intensity-Items in der Kurzform „Pediatric Pain Interference“ enthalten sind, während nur 5 der 8 verwendeten Fatigue-Items und 5 der 10 Mobility-Items Bestandteil der entsprechenden Kurzformen sind.

Basierend auf Wachstumskurvenmodellen (Parallel Process Longitudinal Growth Models) mit Longitudinaldaten von Kindern und Jugendlichen (n = 82) mit chronischen muskuloskeletalen Schmerzen wurde die Kongruenz anhand von Korrelationen der Kurvenverläufe der Pain-Interference-Kurzform zum Functional Disability Inventory – Child Version und der Physical-Function-Mobility-Kurzform zur Physical-Functioning-Skala des PedsQL untersucht. In diesem laut Autorengruppe ersten Validierungsansatz ergaben sich ähnliche Kurvenverläufe, jedoch nicht durchgehend vergleichbare Kurvensteigungen zu den Referenzinstrumenten [19]. Hinweise auf vorhandene Änderungssensitivität der PROMIS-Kurzform-Skalen „Pediatric Pain Interference“, „Physical Function Mobility“ und „Fatigue“ wurde ebenfalls in dieser Patientenpopulation gezeigt. Grundlage war ein Wachstumskurvenmodell mit Daten von 145 Kindern und Jugendlichen mit chronischen Schmerzen (z. B. Rückenschmerzen, Fibromyalgie, komplexe regionale Schmerzen oder verstärkte regionale Schmerzen) [19]. Untersuchungen der Änderungssensitivität anhand eines patientenberichteten Ankers liegen nicht vor.

Instrumente zur Fremdeinschätzung (Elternfragebogen/Proxy-Versionen)

Basierend auf dem initialen pädiatrischen Itempool aus 293 Items wurden Proxy-Versionen der pädiatrischen PROMIS-Itembänke entwickelt, die eine Fremdeinschätzung der Kinder durch ihre Eltern ermöglichen sollen. Zunächst wurden die Items der Selbsteinschätzung durch Umformulierungen in Proxy-Items konvertiert und kognitive Interviews zum Verständnis der Items mit 25 Eltern von Kindern zwischen 5 und 17 Jahren durchgeführt. Anschließend wurden die 293 Items bei Eltern (n = 1.980) gesunder und chronisch kranker Kinder zwischen 5 und 17 Jahren getestet. Dafür wurden die Items auf 9 verschiedene Test-Fragebögen verteilt, wobei jedes Item in 3 Fragebögen enthalten war. Jedes Item wurde von mindestens 428 Eltern beantwortet. Bei einem Großteil der Eltern (n = 1.548) waren die Kinder ≥ 8 Jahre alt; diese Kinder beantworteten parallel die Items der Selbstausfüllerversion [16].

Zur Untersuchung der Proxy-Itembänke wurden zunächst die Daten der Eltern von Kindern ab 8 Jahren sowie die Daten der Selbsteinschätzung dieser Kinder verwendet [36]. Konfirmatorische Faktorenanalysen und IRT-Analysen zur Skalendimensionalität

Endpunktkategorie Morbidität

und lokalen Abhängigkeit sowie IRT-Analysen zum DIF wurden durchgeführt. Bei der finalen Kalibrierung der eindimensionalen Domänen wurden die Skalen der Proxy-Version auf die Metrik der Kinder-Items angepasst. Ebenso wie für die Itembänke zur Selbsteinschätzung wurden eindimensionale Kurzformen für die einzelnen Itembänke der Elternversionen entwickelt, wobei im Ergebnis die Proxy-Kurzformen jeweils ein Teilset der Kinderversionen darstellten.

Zweidimensionale IRT-Modelle wurden eingesetzt, um die Korrelation zwischen Eltern- und Kinderversion zu schätzen. Es zeigte sich eine schwache bis moderate Übereinstimmung zwischen Fremd- und Selbsteinschätzung (Korrelationen zwischen 0,4–0,7). Diese Ergebnisse werden von der Autorengruppe als konsistent mit den Angaben in der Literatur beschrieben, wonach Selbsteinschätzungen der Lebensqualität durch die Kinder und Fremdeinschätzungen durch die Eltern je nach Domäne voneinander abweichen können [36]. Angaben zur Änderungssensitivität der Proxy-Itembänke wurden nicht identifiziert.

Zur Untersuchung der Proxy-Versionen für Eltern von Kindern zwischen 5 und 7 Jahren wurden IRT-Analysen zum DIF durchgeführt, um zu überprüfen, ob die Items bei Eltern von jüngeren Kindern dieselben unbeobachteten Konstrukte auf die gleiche Weise messen wie bei Eltern von älteren Kindern [37]. Dazu wurde das DIF zwischen den Elternantworten der Altersgruppen 5–7 und 8–17 Jahre untersucht. Datengrundlage war die Stichprobe der 1.980 Eltern gesunder und chronisch kranker Kinder zwischen 5 und 17 Jahren (s. o. [16]). Items, die DIF aufwiesen, wurden einem 7-köpfigen Expertenpanel vorgelegt; dieses entschied im Konsens unter Berücksichtigung der DIF-Analysen und der Inhalte der Items, ob die entsprechenden Items von der Proxy-Version für Eltern von Kindern zwischen 5 und 7 Jahren gestrichen wurden oder mit einem Hinweis, diese Items nicht für den Vergleich von jüngeren und älteren Kindern zu verwenden, beibehalten wurden. Im Ergebnis wurden die meisten Items auch in der Proxy-Version für Eltern von jüngeren Kindern berücksichtigt. Zu beachten ist, dass ein Teil der Items aufgrund unzureichender Stichprobengröße nicht in die Analyse einfluss [37].

Zusammenfassung

Schmerzen, Fatigue und Einschränkungen in der körperlichen Funktion sind patientenrelevante Endpunkte in der vorliegenden Indikation. Der pU hat die inhaltliche Geeignetheit der ausgewählten Items nachvollziehbar begründet.

Die pädiatrischen PROMIS-Itembänke werden in der Literatur als geeignete Instrumente zur Erhebung der Endpunkte „Schmerzen“, „Fatigue“ und „Einschränkungen in der körperlichen Funktion“ beschrieben. Die Itembänke zur Fremd- und Selbsteinschätzung wurden mittels IRT kalibriert, weiterführende Untersuchungen der psychometrischen Eigenschaften liegen nur für Kurzformen der Selbsteinschätzung vor. Unter anderem liegen Hinweise auf Änderungssensitivität der Kurzformen vor. Jedoch sind nicht alle der in der Studie eingesetzten Items Teil der Kurzformen. Inwieweit Ergebnisse der Kurzformen auf die in der Studie verwendeten Items übertragbar sind, ist unklar. Ebenso kann keine Aussage zur Änderungssensitivität der Proxy-Items getroffen werden.

Patientenrelevanz

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

Die Instrumente werden als patientenrelevant bewertet.

Zusammenfassung

Die Ergebnisse der PROMIS-Domänen sind im Ergebnisteil trotz der beschriebenen Einschränkungen hinsichtlich Validität und Änderungssensitivität abgebildet. Bei der Ergebnisinterpretation ist die gemeinsame Auswertung der Daten aus Selbst- und Fremdeinschätzung zu beachten. Inwiefern diese Darstellung in der PROMIS-Methodik vorgesehen ist, ist aufgrund fehlender Angaben unklar. Grundsätzlich erscheint hierbei

Endpunktkategorie Morbidität	
	die unterschiedliche Einschätzung von Einschränkungen durch die Eltern oder die betroffenen Kinder problematisch. Dies zeigt sich u. a. in den schwachen bis moderaten Übereinstimmungen zwischen Fremd- und Selbsteinschätzung [36]. Aufgrund dieser Limitationen sind Veränderungen innerhalb eines Studienarms nicht bewertbar. Der Vergleich der Studienarme wird durch eine gleiche Verteilung an Fremd- und Selbstbeurteilungen pro Studienarm jedoch dargestellt. Gleichzeitig ist zu beachten, dass für den Anteil der Fremdbeurteilungen keine Angaben zur Änderungssensitivität vorliegen.
Endpunkt	Schmerzintensität (FPS-R)
Operationalisierung	<p>Die FPS-R ist eine selbstberichtete Skala zur Erhebung der Intensität akuter Schmerzen von Kindern [12]. Die FPS-R stellt Schmerzintensität graphisch auf einer 6-Punkte-Skala dar, indem Gesichter mit Werten von 0, 2, 4, 6, 8 und 10 (0 = kein Schmerz; 10 = schlimmster Schmerz) verwendet werden. Laut Testanleitung [15] wird das Kind aufgefordert, auf ein Gesicht zu zeigen, das am besten die Intensität des Schmerzes im aktuellen Moment abbildet (s. Abbildung 2).</p>  <p><i>Abbildung 2: Faces Pain Scale - Revised, bewertet 0-2-4-6-8-10 [29]</i></p> <p>Laut Studienunterlagen [29,32] wurde die FPS-R bei allen Kindern eingesetzt, die zu Baseline mindestens 5 Jahre alt waren. Abweichende Angaben finden sich im COA-Dossier (Version 5.0) [30], laut dessen der Einsatz ausschließlich in der Altersgruppe der 5- bis 7-Jährigen stattfand. Im Herstellermodul [21] wird ebenfalls betont, dass das Instrument bei allen Kindern ab 5 Jahren zum Einsatz kam.</p> <p>Erhebungszeitpunkte Baseline, Wochen 24, 40, 64 (= Studienende Japan und Südkorea), 88, 112, 140 (oder vorzeitiger Abbruch). Die Auswertung erfolgte auf Basis der FAS-Population (≥ 5-Jährige) für den Datenschnitt zu Woche 64.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nachvollziehbar beschrieben. Durch die Angaben im Herstellermodul wird davon ausgegangen, dass trotz unterschiedlicher Angaben der Fragebogen bei allen Kindern > 5 Jahren vollumfänglich eingesetzt wurde.</p> <p>Validität Die Skala ist das Ergebnis einer Überarbeitung der originalen Faces Pain Scale (FPS) [1], welche Schmerzintensität auf einer 7-Punkte-Skala abbildete. Durch die Überarbeitung sollte eine Vergleichbarkeit mit 0–5- oder 0–10-Antwortformaten anderer vorhandener Schmerzskaalen geschaffen werden [12]. Unter Verwendung einer computeranimierten Version mit 101 Gesichtern der FPS wurden 15 Erwachsene (im Alter von 18 bis 44 Jahren) gebeten, einen Gesichtsausdruck auszuwählen, der den Zwischenwerten 2, 4, 6 oder 8 entsprach. Das erste und das letzte Gesicht der ursprünglichen FPS wurden als äußere Ankerpunkte festgelegt. Im Ergebnis zeigte sich, dass die Unterschiede im Gesichtsausdruck innerhalb 1 SEM minimal waren. Zudem zeigte sich eine hohe Übereinstimmung in der Auswahl der vier finalen Gesichter hinsichtlich Inter-Rater-Korrelationen zwischen 0,84–0,99 [12]. In der zweiten Phase verwendeten 76 Kinder im Alter von 5 bis 12 Jahren die neue FPS-R mit 6 Gesichtern, um die Intensität der Schmerzen beim Ohrstechen zu bewerten. Neben der FPS-R wurde eine visuelle Analogskala (0 = keine Schmerzen,</p>

Endpunktkategorie Morbidität	
	<p>100 = sehr starke Schmerzen) eingesetzt. Es zeigten sich starke Korrelationen in der Bewertung der beiden Instrumente ($r = 0,93$; $p < 0,001$). In der dritten Phase verwendete eine Stichprobe von Kindern im Alter von 4 bis 12 Jahren die FPS-R und randomisiert zugeteilt zusätzlich entweder</p> <ul style="list-style-type: none"> • eine visuelle Analogskala (0 = keine Schmerzen; 10 = schlimmster Schmerz) oder • eine farbige Analogskala (links schmal und weiß = keine Schmerzen; rechts breit und dunkelrot = stärkste Schmerzen). <p>Alle Kinder ($N = 90$) bewerteten Schmerzen während des Krankenhausaufenthaltes bei chirurgischen und nicht-chirurgischen Schmerzzuständen. Der Pearson-Korrelationskoeffizient zwischen der FPS-R und der visuellen bzw. farbigen Analogskala lag in allen Altersgruppen über $r > 0,7$. Keine signifikanten Unterschiede zeigten sich zwischen den Mittelwerten auf der FPS-R und den beiden Analogskalen [12].</p> <p>Untersuchungen der Änderungssensitivität anhand eines patientenberichteten Ankers liegen nicht vor. Eine Studie liefert jedoch Hinweise auf Änderungssensitivität der FPS-R mit 29 Kindern zwischen 9 und 18 Jahren nach einem chirurgischen Eingriff. Der Steigungskoeffizient im linearen Modell war statistisch signifikant und, wie von der Autorengruppe erwartet, in inverser Richtung mit der Prädiktorvariable Zeit korreliert [4].</p> <p>Der pU stellt im COA-Dossier eigene Untersuchungen zum Verständnis der FPS-R bei XLH-Erkrankten dar. Dazu wurden 4 Eltern von an XLH erkrankten Kindern im Alter von 5 bis 7 Jahren befragt. Zwei Eltern gaben an, dass ihre Kinder die Skala verwenden und zwischen den Gesichtern unterscheiden könnten. Eine Mutter äußerte Zweifel daran, dass ihr Kind die vier Gesichter in der Mitte differenzieren könnte und ein Vater gab an, dass sein Kind die Gesichter nicht gut verstehen würde [30].</p> <p>Patientenrelevanz</p> <p>Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p> <p>In der Nutzenbewertung werden daher die Ergebnisse der FPS-R berücksichtigt.</p> <p>Gleichzeitig ist anzumerken, dass die Erhebung aktuell bestehender Schmerzen zu Studienvisiten in einem Anwendungsgebiet mit chronischem Schmerzbild nicht das Spektrum der auftretenden Symptomatik vollumfänglich erfasst.</p>
Endpunkt	Dentale Ereignisse
Operationalisierung	<p>Dentale Ereignisse bilden einen kombinierten Endpunkt, laut Studienprotokoll und SAP definiert als Anzahl zahnmedizinischer Ereignisse bei Zahnkaries, verzögerter Durchbruch der Zähne, Zahnschmelzhypoplasie, Zahnabszess, Zahnextraktion (durch Karies oder Abszess), Wurzelbehandlung und/oder Gingivitis, die zu Studienbeginn und nach Verabreichung der Studienmedikation beurteilt wurden.</p> <p>Im Herstellermodul findet sich eine andere Aufzählung: Hiernach wurden „Karies, verzögertes Zahnen, Zahnschmelzhypoplasie, Abszesse und Gingivitis“ [21] als Ereignisse erfasst. Zahnextraktion und Wurzelbehandlung fehlen in dieser Aufzählung. Im Studienbericht werden wiederum nur die Ereignisse Zahnabszess, Zahnextraktion (durch Karies oder Abszess), Wurzelbehandlung, Zahnkaries und Gingivitis berichtet. Im Vergleich zum Studienprotokoll fehlen verzögerter Durchbruch der Zähne und Zahnschmelzhypoplasien.</p> <p>Die Erhebung des Endpunktes erfolgte als orale Untersuchung („oral examination“) durch das klinische Studienpersonal für das Safety Set zu jeder Studienvisite. Zudem wurden die Kinder befragt, ob eines der oben genannten Ereignisse seit der letzten Studienvisite aufgetreten ist. Im Herstellermodul wird lediglich die Befragung der Kinder zur Erfassung benannt.</p> <p>Das klinische Studienpersonal stufte anschließend ein, „ob es sich um Zahnbeschwerden oder ein UE handelt[e]“ [21]. Laut Studienbericht wurden Ergebnisse der oralen</p>

Endpunktkategorie Morbidität	
	<p>Untersuchung separat von dentalen Sicherheitsereignissen berichtet, da die dentalen Ereignisse proaktiv erfasst wurden, dentale Sicherheitsereignisse hingegen nur, sofern sie vom klinischen Studienpersonal bemerkt wurden. Diese Information zur Kategorisierung von dentalen Ereignissen in Morbidität und Sicherheit findet sich nicht in Studienprotokoll oder SAP.</p> <p>Separate zahnärztliche Beurteilungen sind nicht erfolgt. Es werden keine Schulungen des Studienpersonals durch Erfassung und Kategorisierung der zahnmedizinischen Ereignisse beschrieben.</p> <p>Erhebungszeitpunkte Baseline, Wochen 4, 8, 16, 24, 32, 40, 52, 64, 74 (oder vorzeitiger Abbruch). Die Auswertung erfolgte auf Basis des Safety Sets für den Datenschnitt zu Woche 64.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nicht nachvollziehbar beschrieben:</p> <ul style="list-style-type: none"> • Der Endpunkt ist vom pU als "Zahnmedizinische Untersuchungen" benannt. Es scheinen jedoch weder zahnmedizinische Fachleute an der Erhebung des Endpunkts beteiligt worden zu sein noch Schulungen des Studienpersonals stattgefunden zu haben. Laut Herstellermodul handelt es sich nur um eine Befragung der Kinder während der Studienvisiten. Inwiefern zusätzlich oder stattdessen die Eltern der Kinder befragt wurden, geht aus den Unterlagen nicht hervor. • Die Benennung relevanter Ereignisse unterscheidet sich in den Studienunterlagen. • Grundsätzlich ist eine Separierung von Ereignissen der Morbidität und der Sicherheit begrüßenswert. Da jedoch keine Kriterien beschrieben sind, nach denen zwischen Zahnbeschwerden und UE unterschieden wurde, ist unklar, inwiefern dies gelungen ist. Eine nicht-systematische Kategorisierung erscheint sowohl für diesen Endpunkt als auch für die Sicherheitserhebung problematisch. <p>Validität Da keine zahnmedizinischen Fachleute an der Erhebung beteiligt wurden und keine Schulungen des Studienpersonals beschrieben sind, ist die Validität des Endpunkts unklar. Eine Erfassung über Berichte der Kinder erscheint höchstens eingeschränkt valide. Weitere Einschränkungen der Validität entstehen durch die nicht-systematische Kategorisierung in Ereignisse der Morbidität und der Sicherheit.</p> <p>Patientenrelevanz Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“ Die im Studienbericht dargestellten Ereignisse werden als patientenrelevant erachtet.</p> <p>Zusammenfassung Die nicht nachvollziehbare Operationalisierung und die beschriebenen Einschränkungen in der Validität schränken die Aussagekraft des Endpunkts gravierend ein, sodass der Endpunkt als nicht geeignet zur Darstellung von Morbiditätsereignissen eingestuft wird. Der Endpunkt wird jedoch ergänzend zu den Sicherheitsanalysen dargestellt, da die nicht nachvollziehbare Separierung von Sicherheits- und Morbiditätsereignissen zu einem Informationsverlust führen könnte.</p>

Abkürzungen: 6MWT: 6-Minuten-Gehtest; AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; ATS: American Thoracic Society; BOT-2: Bruininks-Oseretsky Test of Motor Proficiency – 2nd Edition; COA: Clinical Outcomes Assessment; DIF: Differential Item Functioning; FAS: Full Analysis Set; FPS-R: Faces Pain Scale - Revised; IRT: Item Response Theory; MID: Minimal Important Difference; SD: Standardabweichung; PedsQL: Pediatric Quality of Life Inventory; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; UE: Unerwünschte/s Ereignis/se; XLH: X-chromosomale Hypophosphatämie.

Tabelle 7: Beschreibung der Endpunkte der Kategorie Lebensqualität

Endpunktkategorie Lebensqualität	
Endpunkt	Short Form Health Survey-10 for Children (SF-10)
Operationalisierung	<p>Der SF-10 ist ein durch Eltern ausgefüllter Fragebogen zur Beurteilung der körperlichen und psychosozialen Lebensqualität bei gesunden und kranken Kindern.</p> <p>Der Fragebogen besteht aus 9 Items, wobei Item 2 Einschränkungen während zwei unterschiedlicher Aktivitäten abfragt. Items 1, 2a, 2b, 3 und 5 bilden den körperlichen Gesamt-Score (PHS-10), Items 4 und 6 bis 9 den psychischen Gesamt-Score (PSS-10). Höhere Werte stellen eine bessere Lebensqualität dar. Die Items des PHS-10 fragen den allgemeinen Gesundheitszustand, Einschränkungen durch Gesundheitsprobleme und Schmerzen ab. Items des PSS-10-Scores bilden Einschränkungen durch psychische Probleme, Verhaltensauffälligkeiten, Zufriedenheit und Freundschaft ab. Der Bezugsraum umfasst 4 Wochen. Laut Studienunterlagen sollte jeder Versuch unternommen werden, den Fragebogen im Studienverlauf durch den gleichen Elternteil ausfüllen zu lassen [32].</p> <p>Die Skalenwerte wurden so standardisiert, dass eine Punktzahl von 50 dem Durchschnittswert in einer Stichprobe aus gesunden Kindern und einem ergänzenden Sample mit Kindern mit Behinderungen und chronischen Erkrankungen entspricht. Die Auswertung des SF-10 erfolgte entsprechend der 2007 veröffentlichten Methodik [25].</p> <p>Erhebungszeitpunkte Baseline, Wochen 24, 40, 64 (= Studienende Japan und Südkorea), 88, 112, 140 (oder vorzeitiger Abbruch). Die Auswertung erfolgte auf Basis der FAS-Population (≥ 5-Jährige) für den Datenschnitt zu Woche 64.</p>
Bewertung	<p>Operationalisierung Die Operationalisierung ist nachvollziehbar beschrieben.</p> <p>Validität Zur Entwicklung des ursprünglichen SF-10 wurde mindestens ein Item jeder Skala des Child Health Questionnaire - Parent Form 50 (CHQ-PF50) ausgewählt. Ausgenommen hiervon war die Skala „Wirkung der Familie und Eltern“, da diese sich auf die Auswirkungen der Kindesgesundheit auf die gesamte Familie konzentriert, nicht auf die Gesundheit des Kindes selbst. Items, die in Modellberechnungen am besten den CHQ-PF50-Score vorhersagen konnten, wurden anschließend für die finale Version des SF-10 ausgewählt. Dafür wurden die Ergebnisse einer Stichprobe von 369 Eltern/ Erziehungsberechtigten eines Kindes zwischen 5 und 18 Jahren aus einer Querschnitterhebung verwendet (NSFHS-Sample). Zudem wurden aus diesem Studiensample Daten zur Untersuchung der Reliabilität und Validität des ursprünglichen SF-10 genutzt. Die diskriminante Validität des originalen SF-10 wurde in der NSFHS-Stichprobe untersucht. Im Handbuch aus dem Jahr 2007 [25] wird mit Verweis auf eine demnächst erscheinende Veröffentlichung beschrieben, dass der SF-10 zwischen Gruppen mit bekannten Gesundheitsproblemen unterscheiden konnte. Psychometrische Gütekriterien wurden darüber hinaus in einem weiteren Studiensample mit 187 Eltern von an Asthma erkrankten Kindern untersucht (AOMS-Sample). Laut Handbuch konnte anhand SF-10 zwischen Kindern mit verschiedenen Asthma-Schweregraden unterscheiden. Die zitierte Veröffentlichung konnte jedoch nicht aufgefunden werden.</p> <p>Ein Jahr nach Veröffentlichung der Entwicklung und Validierung des ursprünglichen SF-10 erschien eine überarbeitete Version des SF-10. Änderungen betreffen den Austausch von 4 Items zur Einschätzung der globalen Gesundheit, des Selbstwertgefühls und der Zufriedenheit (insgesamt und in Bezug zu Freundschaften). Begründet werden diese Änderungen mit Erfahrungen aus der laufenden Forschung [25]. In den Studienunterlagen der Studie UX023-CL201 wird das Handbuch der finalen Fragebogen-Version als Grundlage der Erhebungen zitiert.</p> <p>Der finale Fragebogen wurde in einem für die USA repräsentativen Studiensample mit 3.563 Kindern zwischen 5 und 17 Jahren validiert. Die Studienteilnehmenden erhielten</p>

Endpunktkategorie Lebensqualität	
	<p>nach dem Zufallsprinzip den Fragebogen auf unterschiedlichem Wege (Post: N = 1.284; Internet: N = 2.279). Alle Items korrelierten signifikant höher mit ihrer zugeordneten Skala als mit der alternativen Skala ($r > 0,4$). Die Korrelationen variierten nicht zwischen den Administrationswegen für die PHS-10 (Post: $r = 0,42-0,65$; Internet: $r = 0,47-0,65$) und für die PSS-10 (Post: $r = 0,58-0,67$; Internet: $r = 0,62-0,72$). Die interne Konsistenz war mit $\alpha > 0,7$ akzeptabel (PHS-10: Gesamte Stichprobe: $\alpha = 0,76$; PSS-10: Gesamte Stichprobe: $\alpha = 0,83$). Weitere Untersuchungen der psychometrischen Güte fehlen. Es liegen keine Untersuchungen zum SF-10 bei an XLH erkrankten Personen vor.</p> <p>Zusammenfassend zeigt sich sowohl für den ursprünglichen als auch den finalen SF-10 ein intransparenter Entwicklungsprozess, der so zu starken Einschränkungen der Inhaltsvalidität führt. Akzeptable interne Konsistenz und Strukturvalidität konnten für den finalen SF-10 gezeigt werden. Weitere Untersuchungen zur Reliabilität und Validität der finalen Version fehlen. Gerade in Anbetracht der Fremdadministration wären Untersuchungen der Inter-Rater-Reliabilität angezeigt.</p> <p>Patientenrelevanz</p> <p>Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer <u>Verbesserung der Lebensqualität</u>.“</p> <p>Erhebungen der gesundheitsbezogenen Lebensqualität werden als ein Schwerpunkt patientenrelevanter Endpunkte erachtet. Durch die nicht nachvollziehbare Darstellung der Fragebogenentwicklung kann jedoch beim SF-10 nicht davon ausgegangen werden, dass tatsächlich für Erkrankte relevante Aspekte der gesundheitsbezogenen Lebensqualität erfasst werden.</p> <p>Ob tatsächliche Veränderungen über das Instrument abgebildet werden können, ist durch fehlende Untersuchungen der Änderungssensitivität offen. Zudem ist die Objektivität der erfassten Merkmale durch die Eltern nicht untersucht. Aus diesen Gründen werden die Ergebnisse in der Nutzenbewertung nicht berücksichtigt.</p>

Abkürzungen: AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; PHS-10: körperlicher Gesamt-Score; PSS-10: psychischer Gesamt-Score; SF-10: Short Form Health Survey-10 for Children; XLH: X-chromosomale Hypophosphatämie.

Tabelle 8: Beschreibung der Endpunkte der Kategorie Sicherheit

Endpunktkategorie Sicherheit	
Endpunkt	Unerwünschte Ereignisse
Operationalisierung	<p>Als UE wurde jedes unerwünschte medizinische Ereignis definiert, das mit dem Gebrauch des Medikaments in Verbindung stand. Ein UE konnte daher jedes nachteilige und unerwünschte Symptom (z. B. auch ein abnormaler Laborbefund) oder jede Erkrankung sein, die zeitlich mit der Verwendung des Arzneimittels verbunden war, unabhängig davon, ob es mit dem Arzneimittel in Zusammenhang stand oder nicht.</p> <p>UE wurden bei allen Visiten und im gesamten Studienverlauf gemäß MedDRA (v18.1.) erhoben. Laut originalem Studienprotokoll wurden alle UE ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung bis zu 12 Wochen nach der letzten Gabe des Studienmedikaments aufgezeichnet. Im Studienprotokoll nach Amendment 1 findet sich die Einschränkung, dass nur Personen nachbeobachtet werden, die nicht an einer anderen Studie mit Burosumab teilnehmen und nicht mit kommerziell verfügbarem Burosumab oder mit Burosumab auf einem anderen Wege behandelt werden. UE würden laut pU jedoch außerhalb der Studie 301 weiterhin entweder nach einem anderen Studienprotokoll oder je nach Auflagen der Zulassung durch das Monitoring der Sicherheit gemeldet [21]. Im SAP findet sich zudem die Einschränkung, dass nur UE unter</p>

Endpunktkategorie Sicherheit

Behandlung (TEAE) ausgewertet werden, d. h. UE, die bei oder nach der ersten Dosis des Prüfprodukts auftraten und nicht bereits vor der ersten Dosisgabe bestanden.

Sofern angemessen, wurden klinisch signifikante Veränderungen gegenüber Baseline-Werten bei Untersuchungen, Vitalparametern, klinischen Laborparametern, Nierensonographie, GFR, ECHO und EKG als UE erfasst.

Durch das klinische Studienpersonal fand eine Zuweisung von dentalen Ereignissen in die Kategorien Morbidität und Sicherheit statt (s.o., Tabelle 6).

Schwerwiegende unerwünschte Ereignisse (SUE)

Ein SUE ist ein unerwünschtes Ereignis, das nach Einschätzung des klinischen Studienpersonals oder des pU folgende Auswirkungen hatte:

- Tod,
- ein lebensbedrohliches UE,
- stationärer Krankenhausaufenthalt oder Verlängerung des bestehenden Krankenhausaufenthaltes,
- anhaltende oder erhebliche Unfähigkeit oder erhebliche Beeinträchtigung der Fähigkeit normale Lebensfunktionen auszuführen,
- eine angeborene Anomalie/Geburtsstörung.

Krankenhausaufenthalte, die vor der Studienteilnahme geplant waren (z. B. für optionale Operationen), wurden nicht als SUE betrachtet. Krankenhausaufenthalte aufgrund von Vorerkrankungen, die nach Eintritt in die Studie geplant wurden, galten als SUE.

Medizinische Ereignisse, die nicht zum Tod führten, nicht sofort lebensbedrohlich waren oder einen Krankenhausaufenthalt erforderten, konnten zudem als schwerwiegend angesehen werden, wenn sie Studienteilnehmende gefährden oder einen medizinischen oder chirurgischen Eingriff erforderten, um eines der oben genannten Ereignisse zu verhindern.

Schweregrad von UE

Soweit möglich, wurde der Schweregrad der UE anhand der CTCAE (v4) des National Cancer Institute (NCI) bewertet.

Wenn UE nicht nach CTCAE bewertet werden konnte, sollten sie als mild, mäßig, schwer, lebensbedrohlich oder Tod unter Verwendung der folgenden Definitionen eingestuft werden:

- Mild (Klasse 1): Wahrnehmung von Symptomen, aber leicht verträglich und wenig störend und verursachen nicht eine Steigerung des zeitlichen Aufwands normaler Aktivitäten. Symptome erfordern keine Therapie oder eine medizinische Beurteilung; Anzeichen und Symptome sind vorübergehend.
- Mäßig (Klasse 2): Symptome führen zu geringen Unannehmlichkeiten oder Sorgen und können tägliche Aktivitäten stören, werden aber in der Regel durch einfache therapeutische Maßnahmen behandelt.
- Schwer (Klasse 3): Symptome unterbrechen den normalen Tagesablauf und führen in der Regel zu einer systemischen medikamentösen Therapie oder einer anderen Behandlung; sie führen in der Regel zu Handlungsunfähigkeit.
- Lebensbedrohlich (Klasse 4): Ereignisse, die den Studienteilnehmer einer unmittelbaren Todesgefahr aussetzen oder handlungsunfähig machen.
- Tod (Klasse 5): Ereignisse, die zum Tode führen.

UE von besonderem Interesse

Im SAP vordefinierte Ereignisse von besonderem Interesse waren:

- Reaktionen an der Injektionsstelle: Definiert durch Preferred Terms (PT) im MedDRA-Oberbegriff (high-level term, HLT) „Reaktion an der Injektionsstelle“.
- Hypersensibilität: Definiert durch relevante PT im standardisierten MedDRA-Query (SMQ) für „Hypersensibilität“.
- Hyperphosphatämie: Definiert durch PT „Hyperphosphatämie“ und „Blutphosphat erhöht“.

Endpunktkategorie Sicherheit	
	<ul style="list-style-type: none"> • Ektopische Mineralisierung: Da laut pU kein SMQ verfügbar ist, wurde das UE mit Hilfe einer MedDRA-Suche von „Mineralisierung“ definiert. • Restless-Legs-Syndrom: Definiert durch die PT „Restless-Legs-Syndrom“, „Restlessness“, „Akathisia“, „Sinnesstörungen“, „Psychomotorische Hyperaktivität“, „Gliederschmerzen“, „Neuromuskuläre Schmerzen“ und „Formation“. <p>Erhebungszeitpunkte</p> <p>Jede Studienvisite. Weitere Hausvisiten waren im Burosumab-Arm je nach Nähe zum Studienzentrum und Patientenpräferenz während der Behandlungsperiode möglich (nicht in Japan und Südkorea). Laut Studienprotokoll wurden im Kontrollarm Sicherheitsereignisse außerhalb der Studienvisiten telefonisch erfragt. In der Extensionsperiode waren Hausvisiten für alle Studienteilnehmenden möglich (nicht in Japan und Südkorea).</p> <p>Für Studienteilnehmende, die die Behandlung nach Woche 64 nicht fortsetzten (z. B. durch Studienende in Japan und Korea) und nicht in einem anderen Rahmen weiterbeobachtet wurden, fand eine weitere Studienvisite nach 12 (\pm 1) Wochen nach der letzten Dosis statt.</p> <p>Die Analysen beruhen auf dem Safety Set. Es liegen getrennte Analysen nach Studienperiode vor.</p>
Bewertung	<p>Operationalisierung</p> <p>Die Operationalisierung ist mit Einschränkungen nachvollziehbar:</p> <ul style="list-style-type: none"> • Es ist unklar, inwiefern die unterschiedliche Erhebung von Sicherheitsereignissen außerhalb der Studienvisiten (im Interventionsarm als Hausvisiten, im Kontrollarm telefonisch) Auswirkungen auf das Berichten von Ereignissen hatte. Es ist zudem unklar, ob Teilnehmende des Interventionsarms ohne Hausvisiten (z. B. in Japan und Südkorea) ebenfalls telefonisch befragt wurden. • Die bereits beim Endpunkt „Dentale Ereignisse“ (s. Tabelle 6) thematisierte, nicht-systematische Kategorisierung in Ereignisse der Morbidität und Sicherheit erscheint auch für die Sicherheitserhebung problematisch. <p>Validität</p> <p>In Studie 301 werden UE ab Unterzeichnung der Einverständniserklärung erfasst, jedoch erst ab der ersten Dosis der jeweiligen Studienmedikation ausgewertet. Es ist unklar, wieso UE erhoben, jedoch nicht ausgewertet werden. Der Endpunkt wird bis Woche 64 dennoch als valide bewertet.</p> <p>Durch die geplante selektive Nachbeobachtung im Rahmen der Studie 301 ist die Validität nach Woche 64 potentiell eingeschränkt. In der Extensionsperiode werden lediglich die Personen nachbeobachtet, die Burosumab im Rahmen der Studie 301 weiter erhielten. Sofern die Teilnehmenden aus der Studie nach Woche 64 ausschieden, was in Japan und Korea regelhaft der Fall war, konnte die Sicherheitserhebung auch außerhalb der Studie stattfinden. Für Studie 301 könnte dadurch eine mögliche Selektion in der Nachbeobachtung stattfinden.</p> <p>Patientenrelevanz</p> <p>Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der <u>Verringerung von Nebenwirkungen</u> oder einer Verbesserung der Lebensqualität.“</p> <p>Die Ergebnisse zum Endpunkt „Unerwünschte Ereignisse“ werden in die Bewertung des Zusatznutzens von Burosumab eingeschlossen. Die Ergebnisse des Endpunkts „Dentale Ereignisse“ werden hier ergänzend dargestellt.</p>

Abkürzungen: AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung; CTCAE: Common Terminology Criteria for Adverse Events; MedDRA: Medical Dictionary for Regulatory Activities; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; SUE: Schwerwiegende/s unerwünschte/s Ereignis/se TEAE: Treatment Emergent Adverse Events; UE: Unerwünschte/s Ereignis/se.

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

- Mortalität,
- Körpergröße,
- Gehfähigkeit mittels 6MWT (6-Minuten-Gehtest),
- PROMIS-Domänen „Pediatric Pain Interference“, „Physical Function Mobility“, „Fatigue“,
- Schmerzintensität mittels FPS-R (Faces Pain Scale - Revised).

Der Endpunkt „Dentale Ereignisse“ wird in dieser Nutzenbewertung ergänzend in der Kategorie Sicherheit dargestellt. Eine Übersicht der Erhebungszeitpunkte der Endpunkte findet sich in Tabelle 9.

Tabelle 9: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der aktiv kontrollierten Behandlungsperiode der Studie 301

Studienvisite (Woche)	W0 ¹⁾	W4	W8	W16	W24	W32	W40	W52	W64/ET
Mortalität ²⁾⁴⁾	x	x	x	x	x	x	x	x	x
Körpergröße	x	-	-	-	x	-	x	-	x
Gehfähigkeit (6MWT)	x	-	-	-	x	-	x	-	x
PROMIS-Domänen: Pediatric Pain Interference Physical Function Mobility Fatigue	x	-	-	-	x	-	x	-	x
Schmerzintensität (FPS-R)	x	-	-	-	x	-	x	-	x
Dentale Ereignisse ³⁾	x	x	x	x	x	x	x	x	x
Sicherheit ⁴⁾	x	x	x	x	x	x	x	x	x

¹⁾ Entspricht Baseline.

²⁾ Im Rahmen der Sicherheitserhebung zu jeder Studienvisite erhoben.

³⁾ Endpunkt wird ergänzend in der Kategorie Sicherheit dargestellt.

⁴⁾ Zu jeder Studienvisite und darüber hinaus je nach Studienarm in Hausvisiten oder telefonisch (s. Tabelle 8) erhoben.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; ET: Early Termination; FPS-R: Faces Pain Scale - Revised; PROMIS: Patient-Reported Outcomes Measurement Information System; W: Woche.

2.4 Statistische Methoden

In Studie 301 wurde die Überlegenheit von Burosumab gegenüber einer Therapie mit Vitamin D und Phosphat im Parallelgruppendesign getestet. Der finale SAP (Version 1.0) vom 2. Januar 2018 wurde vor dem ersten Datenschnitt zu Woche 40 (12. Februar 2018) finalisiert.

Analysepopulationen

Es wurden drei mögliche Analysepopulationen in Studie 301 (s. Tabelle 10) bestimmt. Die Analysen der Wirksamkeitsendpunkte wurden im Full Analysis Set, die der Sicherheitsendpunkte im Safety Set geplant.

Tabelle 10: Analysepopulationen

Population	Definition
Full Analysis Set (FAS)	Alle randomisierten Patientinnen und Patienten mit mindestens einer Dosis der zugeteilten Studienmedikation und mindestens einer Post-Baseline-Bewertung. Die Auswertung erfolgt entsprechend der randomisierten Behandlung.
Full Analysis-Subset ≥ 5 Jahre	Alle Patientinnen und Patienten des FAS mit einem Alter ≥ 5 Jahre zu Baseline (Stratifizierungsfaktor bei Randomisierung). Die Auswertungen der Endpunkte PROMIS, 6MWT und FPS-R erfolgen mit dieser Teilpopulation.
Safety Set	Alle randomisierten Patientinnen und Patienten mit mindestens einer Dosis der zugeteilten Studienmedikation. Die Auswertung erfolgt entsprechend der tatsächlich erhaltenen Behandlung.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; FPS-R: Faces Pain Scale - Revised; PROMIS: Patient-Reported Outcomes Measurement Information System.

Statistische Analysen

Statistische Tests wurden zweiseitig auf dem Signifikanzniveau $\alpha = 0,05$ und mit 95%-Konfidenzintervallen durchgeführt. Eine Adjustierung für multiple Tests wurde nicht durchgeführt. Sofern Annahmen der Normalverteilung als verletzt eingeschätzt wurden, war die Anwendung nicht-parametrischer Verfahren vorgesehen.

Alle Endpunkte mit wiederholten Messungen während der Studie wurden mit Generalized Estimation Equation (GEE-)Modellen zur Bestimmung des Gruppenunterschieds in den mittleren Veränderungen innerhalb der Studienarme über die Zeit hinweg ausgewertet. Dies trifft auf alle in der Nutzenbewertung dargestellten Wirksamkeitsendpunkte (6MWT, PROMIS, FPS-R, Körpergröße) zu. Im Modell für 6MWT, PROMIS und FPS-R war eine Adjustierung für Behandlungsgruppe, Studienvisite, Interaktionsterm aus Behandlungsgruppe und Studienvisite, dem jeweiligen Wert zu Baseline und für die Stratifizierungsfaktoren RSS zu Baseline ($\geq / < 2,5$) und Altersgruppe ($\geq / < 5$ Jahre) vorgesehen. Um die Kovarianzstruktur zu modellieren, wurde zunächst die austauschbare Kovarianzmatrix ausgewählt. Führt diese zu einer Nichtkonvergenz, sollte eine unabhängige Kovarianzstruktur angewendet werden. Hypothesen wurden nicht formuliert.

Stehhöhe und Liegelänge wurden in cm sowie als z-Scores und Perzentile dargestellt, Sitzhöhe hingegen nur in cm. Die Auswertung der Sitzhöhe erfolgte deskriptiv. Für die Veränderung der Stehhöhe und Liegelänge wurde hingegen ein GEE-Modell verwendet. Es wurde für Behandlungsgruppe, RSS-Gesamtwert zu Baseline, Alter zu Baseline und z-Score der Körpergröße zu Baseline adjustiert. Der mittlere z-Score der elterlichen Körpergröße (gemittelt aus beiden Eltern) konnte als optionale Kovariate ebenfalls in Betracht kommen.

Dentale Ereignisse und Sicherheitsereignisse wurden deskriptiv ausgewertet. Im Herstellerdossier werden zusätzlich post hoc Effektschätzer für Sicherheitsereignisse berichtet.

Imputationsregeln

Es wurden keine Imputationsregeln für die in der Nutzenbewertung enthaltenen Endpunkte beschrieben. Sofern nicht anders angegeben, sollten nur beobachtete Daten verwendet werden.

Subgruppenanalysen

Für den Endpunkt Körpergröße waren im SAP Subgruppenanalysen mit den Merkmalen RSS-Gesamtwert ($\leq 2,5$ vs. $> 2,5$), Region (Japan vs. Rest der Welt), Geschlecht (weiblich vs. männlich)

und – sofern angemessen – nach PHEX-Mutationsstatus (positiv vs. nicht-positiv) präspezifiziert. Es sind keine Kriterien bestimmt worden, anhand derer der pU eine Angemessenheit entschied. Für die Endpunkte 6MWT, PROMIS und FPS-R konnten Subgruppenanalysen für das Merkmal RSS-Gesamtwert durchgeführt werden.

Mit dem Herstellerdossier [21] hat der pU Subgruppenanalysen inklusive Interaktionstests für die relevanten Endpunkte Körpergröße, 6MWT, PROMIS, FPS-R und Sicherheit mit den folgenden Subgruppenmerkmalen eingereicht:

- RSS-Gesamtwert ($\leq 2,5$ vs. $> 2,5$),
- Geschlecht (weiblich vs. männlich),
- Alter (< 5 vs. ≥ 5 Jahre),
- Region (Japan vs. Rest der Welt),
- PHEX-Mutationsstatus (positiv vs. nicht-positiv),
- Orthopädische Chirurgie während der Studie (ja vs. nein).

Für den Endpunkt „Körpergröße“ entsprechen diese Analysen (ausgenommen „Alter“ und „Orthopädische Chirurgie während der Studie“) denen des SAP. Für die Endpunkte 6MWT, PROMIS und FPS-R sind hingegen alle Analysen außer nach RSS-Gesamtwert post hoc durchgeführt worden.

Im Ergebnisteil werden statistisch signifikante Interaktionen ($p < 0,05$) berichtet, sofern „[...] jede Subgruppe mindestens 10 Personen umfasst und bei binären Daten mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind“ [14].

2.5 Verzerrungspotential auf Studien- und Endpunktebene

In Tabelle 11 ist das Ergebnis der Bewertung des Verzerrungspotentials auf Studienebene, in Tabelle 12 auf Endpunktebene abgebildet. Die Bewertung bezieht sich auf den Zeitraum bis Ende der Behandlungsperiode zu Woche 64. Ab Woche 64 ist das Studiendesign in der Extensionsperiode einarmig und es entstehen weitere Aspekte im Verzerrungspotential durch die Gestaltung der Nachbeobachtung (s. Tabelle 5 und Tabelle 8), was jedoch nicht Teil der hier dargestellten Bewertung ist.

Tabelle 11: Verzerrungspotential in der Behandlungsperiode der Studie 301

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung Patientin/Patient	Verblindung Studienpersonal	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
UX023-CL301	Ja ¹⁾²⁾	Ja ³⁾	Nein ⁴⁾	Nein ⁴⁾	Nein	Ja ⁵⁾	Hoch

¹⁾ Randomisierungssequenz erstellt durch einen unabhängigen Drittanbieter. Randomisierungsmethode nicht benannt.

²⁾ Randomisierung war stratifiziert nach RSS zu Baseline (RSS-Gesamtwert $\leq 2,5$ vs. $> 2,5$), Alter (< 5 vs. ≥ 5 Jahre) und Region (Japan vs. Rest der Welt).

³⁾ Zuteilung der Randomisierung erfolgte über ein Interactive Web Response System.

⁴⁾ Offenes Studiendesign.

⁵⁾ Es wird angenommen, dass mit Amendment 1 des Studienprotokolls entgegen GCP-Richtlinien nicht ausschließlich prospektive Änderungen für die Studiendurchführung beschrieben wurden, sondern dass bereits stattgefundenen Änderungen in Amendment 1 nachträglich dokumentiert wurden. Wann die beschriebenen Änderungen tatsächlich eintraten, ist damit häufig unklar. Damit ergibt sich eine deutliche Unsicherheit hinsichtlich der Studiendurchführung.

Abkürzungen: GCP: Good Clinical Practice; RSS: Rickets Severity Scale.

Das Verzerrungspotential auf Studienebene wird als hoch eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 12 dargestellt.

Tabelle 12: Verzerrungspotential in der Behandlungsperiode der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie 301

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Mortalität ¹⁾	Nein ²⁾	Ja	Nein	Nein	Niedrig
Körpergröße	Nein ²⁾	Ja	Nein ³⁾	Nein	Niedrig
Gehfähigkeit (6MWT)	Nein	Ja	Unklar ⁴⁾⁵⁾	Ja ⁶⁾	Hoch
PROMIS-Domänen: Pediatric Pain Interference Physical Function Mobility Fatigue	Nein	Ja	Nein	Ja ⁶⁾	Hoch
Schmerzintensität (FPS-R)	Nein	Ja	Nein ⁷⁾	Ja ⁶⁾	Hoch
Dentale Ereignisse	Nein	Ja	Nein	Ja ⁶⁾⁸⁾	Hoch
Sicherheit	Nein	Ja	Nein	Ja ⁶⁾⁸⁾⁹⁾	Hoch

¹⁾ Im Rahmen der Sicherheit erfasst.

²⁾ Der Einfluss der fehlenden Verblindung wird als gering eingestuft.

³⁾ In der Analyse zum z-Score der Stehhöhe/Körperlänge fehlen Daten von einer Person im Interventionsarm. Da die Messung bereits zu Baseline fehlte, ist nicht von einer systematischen Verzerrung auszugehen.

⁴⁾ Laut Studienbericht fehlten für zwei Kinder des Burosumab-Studienarms Daten: Ein Kind absolvierte den 6MWT nicht zu Baseline, ein Kind nicht zu Woche 40. Es ist unklar, wie viele Kinder in der Analyse berücksichtigt wurden. Der pU berichtet im Herstellermodul nicht a priori geplante Analysen ohne die beiden Kinder mit fehlenden Werten.

⁵⁾ Die Analysen waren nicht (wie a priori geplant) für Alter adjustiert.

⁶⁾ Es wurden höhere Raten an Begleitmedikation (s. Tabelle 16) im Interventionsarm im Studienverlauf festgestellt, u. a. Schmerzmittel.

⁷⁾ Es fehlen Daten von einer Person im Kontrollarm in der Analyse. Da die Messung bereits zu Baseline fehlte, ist nicht von einer systematischen Verzerrung auszugehen.

⁸⁾ Es fand eine nicht-systematische Kategorisierung in Ereignisse der Morbidität und Sicherheit statt (s. Tabelle 6).

⁹⁾ Da im Interventionsarm Hausvisiten stattfinden konnten (außer in Japan und Südkorea), ist von einer stärkeren Erfassung von Sicherheitsereignissen im Interventionsarm gegenüber dem Kontrollarm auszugehen. Im Kontrollarm sollten alternativ telefonische Abfragen stattfinden, wobei unklar ist, ob diese im gleichen Abstand und in der gleichen Häufigkeit wie Hausvisiten stattfanden. Zudem ist denkbar, dass das unterschiedliche Setting der Visiten einen Einfluss auf die Erhebung bzw. auf das Berichten von UE hatte. Es ist nicht bekannt, wie viele Personen aus dem Interventionsarm wie viele Hausvisiten in Anspruch nahmen und wie häufig UE telefonisch erfragt wurden.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; FPS-R: Faces Pain Scale - Revised; ITT: Intention-to-Treat; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; UE: Unerwünschte/s Ereignis/se

3 Ergebnisse der eingeschlossenen Studie

3.1 Studiencharakteristika und Studienmedikation

Auf den Kontrollarm der Studie 301 wurden 32 Personen randomisiert zugeordnet, auf den Interventionsarm 29 (s. Tabelle 13). Diese Personen bilden sowohl die FAS- als auch die Safety-Set-Population. Keine Testperson brach die Einnahme der Studienmedikation vorzeitig ab. Zum Zeitpunkt des Datenschnitts hatten alle Studienteilnehmenden die Behandlungsperiode mit Studienwoche 64 abgeschlossen. Im Interventionsarm waren 25 der 29 Studienteilnehmenden in die Extensionsperiode übergegangen. Im Kontrollarm sind 26 Personen in die Extensionsperiode übergegangen, von denen 22 eine Behandlung mit Burosumab begonnen.

Tabelle 13: Allgemeine Angaben zum Patientenfluss in Studie 301 mit Datenschnitt zu Woche 64

UX023-CL301	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Randomisierte Patientinnen und Patienten ¹⁾ Davon Patientinnen und Patienten ≥ 5 Jahre ²⁾	29 (100) 15 (51,7)	32 (100) 20 (62,5)
Safety Set	29 (100)	32 (100)
Abbruch der Einnahme der Studienmedikation	0	0
<i>Studienphasen</i>		
Behandlungsperiode (Studienwoche 64) abgeschlossen	29 (100)	32 (100)
Übergang in einarmige Extensionsperiode ³⁾	25 (86,2) ⁴⁾	26 (81,3) ⁴⁾

¹⁾ Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (s. Tabelle 10), sodass die FAS-Population der ITT-Population entspricht.

²⁾ Bildet das Full Analysis-Subset ≥ 5 Jahre für die Analysen zu PROMIS, 6MWT und FPS-R.

³⁾ In der einarmigen Extensionsperiode erhielten alle Personen Burosumab. Ein Übergang in die Extensionsperiode war für Studienteilnehmende aus Japan und Südkorea nicht möglich.

⁴⁾ Laut Studienbericht schlossen 7 Personen (2 Personen aus dem Burosumab-Arm und 5 Personen aus dem Kontrollarm, vgl. Tabelle 14) aus Japan und Südkorea die Studie zu Woche 64 ab. Diese Personen waren nicht berechtigt, an der Extensionsperiode teilzunehmen. Drei weitere Testpersonen (2 Personen aus dem Burosumab-Arm und 1 Person aus dem Kontrollarm) entschieden sich für kommerziell erhältliches Burosumab anstelle einer Teilnahme an der Extensionsperiode. Somit traten insgesamt 26 Personen der Kontrollgruppe und 25 Personen der Interventionsgruppe in die Extensionsperiode ein. 22 der 26 Personen aus der Kontrollgruppe hatten zum Zeitpunkt des Datenschnitts mindestens eine Dosis Burosumab erhalten.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; FAS: Full Analysis Set; FPS-R: Faces Pain Scale - Revised; ITT: Intention-to-Treat; PROMIS: Patient-Reported Outcomes Measurement Information System.

Charakteristika der Studienpopulation

Die Charakteristika der Studienpopulation zu Baseline sind in Tabelle 14 dargestellt. Zwischen den Studienarmen zeigen sich weitestgehend gleichmäßige Verteilungen auch außerhalb der Stratifizierungsmerkmale. Auffällig sind in beiden Studienarmen die Perzentile der Stehhöhe/Liegelänge, welche sich in Mittelwert und Median unterscheiden. Durch die Transformation der Werte zeigen sich diese Unterschiede dabei nicht im z-Score.

Tabelle 14: Charakterisierung der FAS-Studienpopulation¹⁾ der Studie 301 zu Baseline

UX023-CL301	Burosumab N = 29	Kontrolle N = 32
<i>Alter (Jahre)</i> MW (SD) Median (min; max) < 5 ²⁾ ≥ 5 ²⁾	5,8 (3,4) 5,8 (0,8; 12,8) 14 (48,3) 15 (51,7)	6,3 (3,2) 6,2 (1,2; 11,9) 12 (37,5) 20 (62,5)
<i>Geschlecht, n (%)</i> männlich weiblich	13 (44,8) 16 (55,2)	14 (43,8) 18 (56,3)
<i>Region, (%)</i> Japan ²⁾ Rest der Welt: ²⁾ Südkorea ³⁾ USA ³⁾ Kanada ³⁾ Schweden ³⁾ Großbritannien ³⁾ Australien ³⁾	2 (6,9) 27 (93,1) 0 16 (55,2) 2 (6,9) 0 2 (6,9) 7 (24,1)	3 (9,4) 29 (90,6) 2 (6,3) 15 (46,9) 7 (21,9) 1 (3,1) 2 (6,3) 2 (6,3)
<i>Ethnie (genetisch), n (%)</i> asiatisch kaukasisch/weiß andere	2 (6,9) 25 (86,2) 2 (6,9)	6 (18,8) 25 (78,1) 1 (3,1)
<i>BMI (kg/m²)</i> MW (SD) Median (min; max)	(n = 28) 18,0 (2,5) 17,5 (14,7; 26,2)	18,2 (2,2) 17,5 (15,5; 26,1)
<i>Stehhöhe/Liege­länge (Perzentile für Alter und Geschlecht)</i> MW (SD) Median (min; max)	(n = 28) 5,9 (10,0) 1,1 (0,0; 40,1)	5,7 (9,5) 1,7 (0,0; 45,0)
<i>Stehhöhe/Liege­länge (z-Score)</i> MW (SD) Median (min; max)	(n = 28) -2,3 (1,2) -2,3 (-5,0; -0,3)	-2,1 (0,9) -2,1 (-4,7; -0,1)
<i>RSS-Gesamtwert zu Baseline</i> MW (SD) Median (min; max) ≤ 2,5 ²⁾ , n (%) > 2,5 ²⁾ , n (%)	3,2 (1,0) 3,0 (2,0; 6,5) 10 (34,5) 19 (65,5)	3,2 (1,1) 3,0 (2,0; 6,5) 12 (37,5) 20 (62,5)
<i>Nierenu­ltraschall-Wert (Skala 0–5), n (%)</i> 0 1 2 3	24 (82,8) 2 (6,9) 2 (6,9) 1 (3,4)	23 (71,9) 3 (9,4) 3 (9,4) 3 (9,4)
<i>Serumphosphat zu Baseline (mg/dl)</i> MW (SD) Median (min; max)	2,4 (0,2) 2,4 (2,0; 2,9)	2,3 (0,3) 2,3 (1,8; 2,8)
<i>PHEX-Mutationsstatus, n (%)</i> positiv negativ wahrscheinlich pathogen un­schlüssige Variante	27 (93,1) 0 2 (6,9) 0	28 (87,5) 1 (3,1) 1 (3,1) 2 (6,3)

UX023-CL301	Burosumab N = 29	Kontrolle N = 32
Medizinische Anamnese		
Jeglicher Befund ⁴⁾⁵⁾ , n (%)	26 (89,7)	26 (81,3)
Muskel- und Bindegewebserkrankungen	17 (58,6)	15 (46,9)
Gastrointestinale Störungen	11 (37,9)	10 (31,3)
Angeborene, familiäre und genetische Störungen	10 (34,5)	9 (28,1)
Allgemeine Erkrankungen u. Beschwerden am Verabreichungsort	9 (31,0)	8 (25,0)
Infektionen und Befall	7 (24,1)	10 (31,3)
Erkrankungen Atemwege, Brustraum und Mediastinum	9 (31,0)	7 (21,9)
Chirurgische und medizinische Prozeduren	10 (34,5)	6 (18,8)
XLH-bezogene Befunde⁴⁾, n (%)		
Biegung der Unterschenkel	24 (82,8)	27 (84,4)
Biegung der Oberschenkel	23 (79,3)	26 (81,3)
Einwärtsgang	14 (48,3)	21 (65,6)
Zahnabszesse	10 (34,5)	11 (34,4)
X-Bein-Stellung	8 (27,6)	6 (18,8)
Gelenksteifigkeit (eingeschränkter Bewegungsumfang)	8 (27,6)	5 (15,6)
Schädelsynostose	4 (13,8)	6 (18,8)
Übermäßiges Karies	2 (6,9)	4 (12,5)
Chiari Malformation	1 (3,4)	1 (3,1)
Vortherapien		
<i>Kinder mit konventioneller XLH-Therapie⁶⁾ vor Studienbeginn, n (%)</i>		
ja	29 (100)	32 (100)
nein	0	0
<i>Dauer der konventionellen XLH-Therapie⁶⁾ (Jahre)</i>		
MW (SD)	3,3 (3,1)	4,3 (3,0)
Median (min; max)	2,2 (0,5; 12,2)	3,5 (0,8; 12,0)
<i>Alter bei Beginn der konventionellen XLH-Therapie⁶⁾ (Jahre)</i>		
MW (SD)	2,6 (2,6)	2,1 (2,0)
Median (min; max)	1,9 (0,1; 11,3)	1,5 (0,0; 6,5)

¹⁾ Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (s. Tabelle 10), sodass die FAS-Population der ITT-Population entspricht.

²⁾ Stratifizierungsmerkmal.

³⁾ Bilden Stratifizierungskategorie „Rest der Welt“.

⁴⁾ Keine Angabe zum Bezugszeitraum im Studienbericht.

⁵⁾ Dargestellt werden Systemorganklassen (MedDRA, Version 18.1) > 30 % in einem Studienarm.

⁶⁾ Konventionelle Therapie umfasste Alfacalcidol, Calcitriol, Calcifediol, K-Phos Neutral, Neutra-Phos(-K), Phos-Nak, Phosphoneurol, Phosphor, Polyfusor Phosphat, Natriumphosphat, zweibasiches Natriumphosphat.

Abkürzungen: FAS: Full Analysis Set; ITT: Intention-to-Treat; MedDRA: Medical Dictionary for Regulatory Activities; MW: Mittelwert; PHEX: Phosphate regulating endopeptidase homolog X-linked; RSS: Rickets Severity Scale; SD: Standardabweichung; XLH: X-chromosomale Hypophosphatämie.

Exposition mit der Studienmedikation

In Tabelle 15 sind Angaben zur Exposition mit der jeweiligen Studienmedikation dargestellt. Mit Datenschnitt zu Woche 64 war die Behandlungsperiode für alle Personen abgeschlossen. Die mediane Behandlungsdauer beträgt daher in beiden Studienarmen 15,2 Monate. Da kein Kind die Behandlung abbrach, entspricht die Behandlungsdauer der Beobachtungsdauer. 22 Kinder aus dem Kontrollarm waren zum Zeitpunkt des Datenschnitts in der einarmigen Extensionsperiode bereits mit Burosumab behandelt worden. Ihre mediane Exposition betrug 1,5 Monate, welche in etwa der Differenz in der Zeitspanne zwischen Ende der Behandlungsperiode und Datenschnitt (1,7 Monate) entspricht.

Tabelle 15: Exposition mit der Studienmedikation in Studie 301

UX023-CL301	Burosumab N = 29	Kontrolle N = 32
<i>Dauer der Exposition, Monate (min; max)</i>		
Mediane Behandlungsdauer in der Behandlungsperiode	15,2 (14,7; 15,5) (n = 29) ¹⁾	15,2 (14,7; 15,3) (n = 22) ¹⁾
Mediane Behandlungsdauer mit Burosumab bis Datenschnitt	17,0 (15,0; 21,3) ²⁾	1,5 (0,2; 5,9) ³⁾
<i>Burosumab-Dosis während der Behandlungsperiode, n (%)</i>		
Beibehaltung der Initialdosis von 0,8 mg/kg	21 (72,4)	n. z. ⁴⁾
Dosissteigerung auf 1,2 mg/kg	8 (27,6)	n. z. ⁴⁾
Burosumab-Dosis in der Extensionsperiode	k. A.	k. A. ⁴⁾
<i>Anzahl der verabreichten Burosumab-Dosen bis Datenschnitt⁴⁾⁵⁾, n (%)</i>		
1–4 Dosen	0	13 (40,6)
10–13 Dosen	0	9 (28,1)
32–45 Dosen	27 (93,1)	0
54 Dosen	1 (3,4)	0
93 Dosen	1 (3,4)	0
<i>Durchschnittliche tägliche Dosis an oralem Phosphat von Baseline zu Woche 64 (mg/kg)</i>		
MW (SD)	n. z.	40,7 (19,6)
Median (min; max)		35,2 (17,8; 101,3)
<i>Durchschnittliche tägliche Dosis an Calcitriol von Baseline zu Woche 64 (mg/kg)</i>		
MW (SD)	n. z.	(n = 22) 0,03 (0,01)
Median (min; max)		0,02 (0,01; 0,05)
<i>Durchschnittliche tägliche Dosis an Alfacalcidol von Baseline zu Woche 64 (mg/kg)</i>		
MW (SD)	n. z.	(n = 9) 0,08 (0,06)
Median (min; max)		0,07 (0,02; 0,22)
<i>Tage mit fehlender Phosphat- / aktiver Vitamin-D-Behandlung bis Ende der Behandlungsperiode, n (%)</i>		
<u>Phosphat</u>		
0		21 (65,6)
1		4 (12,5)
2		4 (12,5)
5	n. z.	2 (6,3)
7		1 (3,1)
<u>Vitamin D</u>		
0		21 (65,6)
1		5 (15,6)
2		4 (12,5)
6		2 (6,3)

¹⁾ Laut Studienbericht schlossen 7 Personen (2 aus dem Burosumab- und 5 aus dem Kontrollarm) aus Japan und Südkorea die Studie zu Woche 64 ab; diese Personen waren nicht berechtigt, an der Extensionsperiode teilzunehmen. 3 weitere Personen (2 aus dem Burosumab-Arm und 1 aus dem Kontrollarm) entschieden sich für kommerziell erhältliches Burosumab anstelle einer Teilnahme an der Extensionsperiode. Somit traten insgesamt 26 Personen der Kontrollgruppe und 25 Personen der Interventionsgruppe in die Extensionsperiode ein. 22 der 26 Personen aus der Kontrollgruppe hatten zum Zeitpunkt des Datenschnitts mindestens eine Dosis Burosumab erhalten.

²⁾ Angabe umfasst Zeit in Behandlungsperiode und Extensionsperiode.

³⁾ Angabe umfasst ausschließlich Zeit in der einarmigen Extensionsperiode.

⁴⁾ Während der kontrollierten Behandlungsperiode erhielt keine Person aus der Kontrollgruppe Burosumab. Im Rahmen der anschließenden einarmigen Extensionsperiode können auch Personen der Kontrollgruppe mit Burosumab behandelt werden. Zum aktuellen Datenschnitt gingen 26 Personen aus dem Kontrollarm in die Extensionsperiode über, von denen 22 mindestens eine Dosis Burosumab erhalten hatten.

⁵⁾ 1. bis 13. Dosis = Dosen in der Extensionsperiode für den Kontrollarm; 32. Dosis und darauffolgende Dosen = Dosen in der Extensionsperiode im Interventionsarm.

Abkürzungen: k. A.: keine Angabe; MW: Mittelwert; n. z.: nicht zutreffend; SD: Standardabweichung.

Begleitmedikation

Die Begleitmedikation in Studie 301 ist in Tabelle 16 dargestellt. Häufigste Begleitmedikationen in beiden Studienarmen waren Analgetika sowie entzündungshemmende und antirheumatische Produkte. Bei 22 der 24 dargestellten Arzneimittelklassen wurden mehr Arzneimittel im Interventionsarm als im Kontrollarm verabreicht. Die größten Unterschiede traten zwischen den Studienarmen bei den Arzneimittelklassen Husten- und Erkältungsmittel (Unterschied: 32%-Punkte), Antibiotika zur systematischen Anwendung (30,5%-Punkte) und Mittel bei obstruktiven Atemwegserkrankungen (25,1%-Punkte) auf.

Tabelle 16: Begleitmedikationen zum Datenschnitt zu Woche 64 bei > 10 % der Patientinnen und Patienten in einer der Behandlungsgruppen der Studie 301

Arzneimittelklasse Arzneimittel	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Analgetika Paracetamol	20 (69,0) 19 (65,5)	19 (59,4) 17 (53,1)
Entzündungshemmende und antirheumatische Produkte Ibuprofen	20 (69,0) 18 (62,1)	18 (56,3) 18 (56,3)
Alle übrigen nichttherapeutischen Mittel	17 (58,6)	13 (40,6)
Antibiotika zur systematischen Anwendung Amoxicillin Azithromycin	17 (58,6) 12 (41,4) 3 (10,3)	9 (28,1) 4 (12,5) 3 (9,4)
Antihistaminika zur systematischen Anwendung Diphenhydramin Loratadin Fexofenadin	13 (44,8) 5 (17,2) 6 (20,7) 3 (10,3)	8 (25,0) 4 (12,5) 2 (6,3) 1 (3,1)
Vitamine Colecalciferol Vitamin D, nicht weiter spezifiziert Vitamine, nicht weiter spezifiziert	13 (44,8) 5 (17,2) 4 (13,8) 3 (10,3)	8 (25,0) 7 (21,9) 2 (6,3) 2 (6,3)
Husten- und Erkältungsmittel Andere Antitussiva	12 (41,4) 4 (13,8)	3 (9,4) 0
Mittel bei obstruktiven Atemwegserkrankungen Salbutamol	10 (34,5) 5 (17,2)	3 (9,4) 2 (6,3)
Anästhetika	7 (24,1)	5 (15,6)
Impfstoffe Influenza-Impfstoffe Masern, Kombinationen mit Mumps und Röteln (lebend abgeschwächt) Varicella Zoster Impfstoffe Poliomyelitis-Impfstoffe Impfstoff Diphtherie-Tetanus-Pertussis	7 (24,1) 3 (10,3) 3 (10,3) 4 (13,8) 3 (10,3) 3 (10,3)	5 (15,6) 3 (9,4) 2 (6,3) 1 (3,1) 1 (3,1) 1 (3,1)
Mittel gegen Obstipation Macrogol	7 (24,1) 3 (10,3)	1 (3,1) 0
Antiemetikum und Mittel gegen Übelkeit Ondansetron	3 (10,3) 3 (10,3)	4 (12,5) 4 (12,5)
Rhinologika	5 (17,2)	2 (6,3)

Arzneimittelklasse Arzneimittel	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Stomatologika	4 (13,8)	3 (9,4)
Antidiarrhoika und intestinale Antiphlogistika/Antiinfektiva Probiotika, nicht weiter spezifiziert	5 (17,2) 3 (10,3)	1 (3,1) 0
Blutersatzmittel und Perfusionslösungen	3 (10,3)	3 (9,4)
Corticosteroide zur systemischen Anwendung	4 (13,8)	2 (6,3)
Otologika	2 (6,9)	4 (12,5)
Psychoanaleptika	3 (10,3)	3 (9,4)
Mineralstoffe	4 (13,8)	1 (3,1)
Ophthalmika	3 (10,3)	2 (6,3)
Alle übrigen therapeutischen Mittel	4 (13,8)	0
Antibiotika und Chemotherapeutika zur dermatologischen Anwendung	3 (10,3)	0
Corticosteroide, dermatologische Zubereitungen	3 (10,3)	0

¹⁾ Entsprechend der Kategorisierungen des WHODRUG 2015Q3.

Protokollverletzungen

Bis zum Datenschnitt zu Woche 64 trat in Studie 301 bei 62,1 % der Kinder im Interventionsarm und bei 40,6 % der Kinder im Kontrollarm eine wesentliche Protokollverletzung auf (s. Tabelle 17). Im Interventionsarm betrifft dies hauptsächlich nicht durchgeführte Untersuchungen und Protokollverletzungen der Kategorie „Andere“, im Kontrollarm sind Protokollverletzungen der Kategorie „Andere“ am häufigsten.

Tabelle 17: Wesentliche Protokollverletzungen in Studie 301 bis zum Datenschnitt zu Woche 64

UX023-CL301	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Jede wesentliche Protokollverletzung ¹⁾	18 (62,1)	13 (40,6)
Untersuchung nicht durchgeführt	9 (31,0)	3 (9,4)
Fehler in der Dokumentation oder verspätete Vervollständigung der Einwilligungserklärung	3 (10,3)	4 (12,5)
Untersuchung nicht im geforderten Zeitraum durchgeführt	1 (3,4)	1 (3,1)
Fehler bei der Dosierung des Studienmedikaments	2 (6,9)	0
Andere ²⁾	8 (27,6)	8 (25,0) ³⁾

¹⁾ Es konnte mehr als eine Protokollverletzung pro Person vorliegen.

²⁾ Betrifft laut Studienbericht hauptsächlich zusätzlich angefertigte Röntgenbilder. Es sind keine konkreteren Informationen verfügbar.

³⁾ Eine fehlende Laboruntersuchung einer Person wurde laut Studienbericht fälschlicherweise hierzu gezählt. In folgenden Berichten soll dies nicht mehr unter „wesentliche Protokollverletzungen“ fallen.

3.2 Mortalität

Es sind keine Todesfälle in Studie 301 aufgetreten.

3.3 Morbidität

Körpergröße

Veränderungen in der Körpergröße, erhoben als Stehhöhe/Liege­länge und Sitzhöhe, bis Woche 64 sind in Tabelle 18 dargestellt. In den z-Scores der Stehhöhe/Liege­länge zeigte sich ein statistisch signifikanter Unterschied zwischen den Studienarmen um 0,1 (95%-KI: [0,0; 0,3]) zugunsten der Intervention. Die klinische Relevanz dieser Differenz bleibt jedoch unklar, da Hedges' g nicht vollständig außerhalb des Irrelevanzbereichs von |0,2| liegt.

Veränderungen in den Perzentilen der Stehhöhe/Liege­länge und in der absoluten Sitzhöhe wurden deskriptiv ausgewertet, sodass Unterschiede zwischen den Studienarmen nicht bewertbar sind.

Tabelle 18: Veränderungen in der Körpergröße zu Woche 64 in Studie 301, FAS-Population¹⁾

UX023-CL301	Burosumab N = 29	Kontrolle N = 32
z-Score der Stehhöhe/Liege­länge		
<i>z-Score zu Baseline</i> MW (SD)	(n = 28) -2,3 (1,2)	-2,1 (0,9)
<i>Veränderung im z-Score zu Woche 64</i> LS Mean (SE) ²⁾³⁾	(n = 28) 0,2 (0,1)	0,0 (0,0)
LS-Mean-Differenz [95%-KI]; p-Wert ²⁾³⁾	0,1 [0,0; 0,3]; 0,0490	
Hedges' g [95%-KI] ⁴⁾	0,5 [-0,03; 1,0]	
Perzentile der Stehhöhe/Liege­länge		
<i>Perzentile zu Baseline</i> MW (SD) Median (min; max)	(n = 28) 5,9 (10,0) 1,1 (0,0; 40,1)	5,7 (9,5) 1,7 (0,0; 45,0)
<i>Veränderung in den Perzentilen zu Woche 64</i> MW (SD) Median (min; max)	(n = 28) 1,5 (6,4) 0,2 (-11,7; 24,6)	-0,7 (3,0) 0,0 (-9,9; 4,4)
Sitzhöhe		
<i>Sitzhöhe in cm zu Baseline</i> MW (SD) Median (min; max)	(n = 26) 59,3 (10,4) 58,3 (27,4; 78,1)	(n = 29) 63,0 (8,6) 61,8 (46,4; 78,4)
<i>Veränderung in der Sitzhöhe in cm zu Woche 64</i> MW (SD) Median (min; max)	(n = 26) 5,1 (5,1) 3,5 (0,8; 24,2)	(n = 29) 3,7 (1,7) 3,8 (1,4; 7,9)

¹⁾ Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (s. Tabelle 10), sodass die FAS-Population der ITT-Population entspricht.

²⁾ GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und Baseline-Alter und Baseline-Stehhöhe/-Liege­länge (z-Score) als kontinuierliche Kovariaten.

³⁾ Für 1 Kind im Burosumab-Arm war keine Angabe zu Baseline verfügbar. Laut SAP wurden fehlende Werte als solche behandelt. Im Herstellermodul wird diese Aussage spezifiziert: „Wenn eine Änderung gegenüber der Baseline vorgenommen wurde, werden nur Patienten mit einer Baseline und mindestens einer Post-Baseline-Messung in die Analyse einbezogen“ [21].

⁴⁾ Angabe aus dem Herstellermodul entnommen, da nicht verfügbar im Studienbericht.

Abkürzungen: FAS: Full Analysis Set; GEE: Generalized Estimation Equation; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; RSS: Rickets Severity Scale SAP: Statistischer Analyseplan; SD: Standardabweichung; SE: Standardfehler.

Gefähigkeit (6MWT)

Die Gefähigkeit wurde bei Kindern mittels 6MWT erhoben, sofern sie zu Baseline mindestens 5 Jahre alt waren. Ergebnisse dieser Erhebung sind in Tabelle 19 dargestellt. Verglichen mit der Kontrollgruppe zeigte sich für die Kinder der Interventionsgruppe zu Woche 64 eine statistisch signifikant größere Veränderung um 43,2 Meter (95%-KI: [2,3; 84,1]). Da Hedges' g nicht vollständig außerhalb des Irrelevanzbereichs von |0,2| liegt, ist die klinische Relevanz dieser Differenz unklar. Im Anteil der erwarteten Laufstrecke zeigte sich ein numerischer Vorteil für die Interventionsgruppe von 6,2% gegenüber der Kontrollgruppe, der jedoch nicht statistisch signifikant ist.

Tabelle 19: Veränderung im 6MWT zu Woche 64 in Studie 301, FAS-Population¹⁾ ≥ 5 Jahre

UX023-CL301	Burosumab N = 15 ²⁾	Kontrolle N = 20 ²⁾
6MWT-Distanz zu Baseline (Meter) MW (SD)	365,9 (118,1)	450,5 (106,4)
Veränderung der 6MWT-Distanz zu Woche 64 (Meter) LS Mean (SE) ³⁾⁴⁾	79,0 (11,4)	35,8 (16,7)
LS-Mean-Differenz [95%-KI]; p-Wert ³⁾⁴⁾⁵⁾	43,2 [2,3; 84,1]; 0,0383	
Hedges' g [95%-KI] ⁶⁾	0,8 [0,1; 1,5]	
Prozentualer Anteil der erwarteten 6MWT-Distanz zu Baseline MW (SD)	62,1 (18,6)	76,2 (14,8)
Veränderung des prozentualen Anteils der erwarteten 6MWT-Distanz zu Woche 64 LS Mean (SE) ³⁾	9,6 (1,9)	3,4 (2,8)
LS-Mean-Differenz [95%-KI]; p-Wert ³⁾⁴⁾⁵⁾	6,2 [-0,7; 13,2]; 0,0781	

¹⁾ Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (s. Tabelle 10), sodass die FAS-Population der ITT-Population entspricht.

²⁾ Der 6MWT wurde ausschließlich bei Kindern durchgeführt, die zum Zeitpunkt der Screening-Visite mind. 5 Jahre alt waren.

³⁾ GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und 6MWT zu Baseline als kontinuierliche Kovariate.

⁴⁾ Laut Studienbericht fehlten für 2 Kinder des Burosumab-Arms Daten: Ein Kind absolvierte den 6MWT nicht zu Baseline, ein Kind nicht zu Woche 40. Laut SAP wurden fehlende Werte als solche behandelt. Im Herstellermodul wird diese Aussage spezifiziert: „Wenn eine Änderung gegenüber der Baseline vorgenommen wurde, werden nur Patienten mit einer Baseline und mindestens einer Post-Baseline-Messung in die Analyse einbezogen“ [21]. Dies steht im Widerspruch zu den Angaben in den statistischen Outputs, nach denen 15 Personen im Burosumab-Arm in die Analysen gingen. Damit ist unklar, wie viele Personen in der Analyse berücksichtigt wurden.

⁵⁾ Im Studienbericht ist anstelle der präspezifizierten FAS-Analyse zu Woche 64 eine nicht a priori vorgesehene Analyse mit Ausschluss der beiden Kinder mit fehlenden Daten enthalten. Diese zeigte Ergebnisse in der gleichen Richtung und vergleichbaren Ausmaßes.

⁶⁾ Angabe aus dem Herstellermodul entnommen, da nicht verfügbar im Studienbericht.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; FAS: Full Analysis Set; GEE: Generalized Estimation Equation; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; RSS: Rickets Severity Scale SAP: Statistischer Analyseplan; SD: Standardabweichung; SE: Standardfehler.

Beeinträchtigung durch Schmerzen, körperliche Funktion und Fatigue: PROMIS

Es wurden bei Kindern, die zu Baseline mindestens 5 Jahre alt waren, drei PROMIS-Domänen eingesetzt. Die Ergebnisse sind in Tabelle 20 abgebildet. In den Domänen zu Beeinträchtigung durch Schmerzen, körperlicher Funktion und Fatigue zeigten sich zu Woche 64 numerische Unterschiede zwischen den Studienarmen, die jedoch keine statistische Signifikanz erreichten.

Tabelle 20: PROMIS-Scores zu Woche 64 in Studie 301, FAS-Population¹⁾ ≥ 5 Jahre

UX023-CL301	Burosumab N = 15 ²⁾	Kontrolle N = 20 ²⁾
Domäne: Paediatric Pain Interference		
Baseline T-Score ³⁾⁴⁾ MW (SD)	53,1 (11,0)	49,9 (12,1)
Veränderung im T-Score zu Woche 64 LS Mean (SE) ⁵⁾	-3,6 (1,9)	-1,3 (1,3)
LS-Mean-Differenz [95%-KI]; p-Wert ⁵⁾	-2,3 [-6,6; 2,1]; 0,3091	
Domäne: Physical Function Mobility		
Baseline T-Score ⁴⁾⁶⁾ MW (SD)	45,2 (9,1)	45,5 (9,9)
Veränderung im T-Score zu Woche 64 LS Mean (SE) ⁵⁾	2,8 (1,6)	0,9 (1,0)
LS-Mean-Differenz [95%-KI]; p-Wert ⁴⁾	1,9 [-1,8; 5,6]; 0,3145	
Domäne: Fatigue		
Baseline T-Score ⁴⁾⁷⁾ MW (SD)	48,8 (9,6)	47,0 (13,7)
Veränderung im T-Score zu Woche 64 LS Mean (SE) ⁵⁾	-3,7 (2,1)	-2,6 (1,5)
LS-Mean-Differenz [95%-KI]; p-Wert ⁵⁾	-1,1 [-6,2; 4,1]; 0,6810	

¹⁾ Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (s. Tabelle 10), sodass die FAS-Population der ITT-Population entspricht.

²⁾ Der PROMIS-Fragebogen wurde ausschließlich bei Kindern eingesetzt, die zum Zeitpunkt der Screening-Visite mindestens 5 Jahre alt waren.

³⁾ Höhere Werte sind gleichbedeutend mit einer stärkeren Beeinträchtigung durch Schmerzen.

⁴⁾ T-Score = standardisierter Score mit Mittelwert von 50 und Standardabweichung von 10.

⁵⁾ GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und Baseline-PROMIS-Domänen-Score als kontinuierliche Kovariate.

⁶⁾ Höhere Werte sind gleichbedeutend mit einer besseren körperlichen Aktivität.

⁷⁾ Höhere Werte sind gleichbedeutend mit einer stärkeren Fatigue.

Abkürzungen: FAS: Full Analysis Set; GEE: Generalized Estimation Equation; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PROMIS: Patient-Reported Outcomes Measurement Information System; RSS: Rickets Severity Scale; SD: Standardabweichung; SE: Standardfehler.

Schmerzintensität (FPS-R)

Kinder, die zu Baseline mindestens 5 Jahre alt waren, verwendeten im Verlauf der Studie 301 die FPS-R zur Bewertung ihrer akuten Schmerzintensität. Bei einer Skala von 0 bis 10 zeigten die Kinder in beiden Studienarmen bei niedrigen Ausgangswerten im Mittel keine Veränderungen bis Woche 64 (s. Tabelle 21).

Tabelle 21: Veränderung im FPS-R zu Woche 64 in Studie 301, FAS-Population¹⁾ ≥ 5 Jahre

UX023-CL301	Burosumab N = 15 ²⁾	Kontrolle N = 20 ²⁾
FPS-R-Wert zu Baseline ³⁾ MW (SD)	0,4 (1,1)	(n = 19) 0,6 (1,2)
Veränderung des FPS-R-Werts zu Woche 64 LS Mean (SE) ⁴⁾⁵⁾	0,0 (0,3)	(n = 19) 0,0 (0,2)
LS-Mean-Differenz [95%-KI]; p-Wert ⁴⁾⁵⁾	0,1 [-0,6; 0,7]; 0,8786	

¹⁾ Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (s. Tabelle 10), sodass die FAS-Population der ITT-Population entspricht.

²⁾ Die FPS-R wurde ausschließlich bei Kindern eingesetzt, die zum Zeitpunkt der Screening-Visite mindestens 5 Jahre alt waren.

³⁾ Höhere Werte sind gleichbedeutend mit einer stärkeren Schmerzintensität, Skala reicht von 0 bis 10.

⁴⁾ Für ein Kind im Kontrollarm war keine Angabe zu Baseline verfügbar. Laut SAP wurden fehlende Werte als solche behandelt. Im Herstellermodul wird diese Aussage spezifiziert: „Wenn eine Änderung gegenüber der Baseline vorgenommen wurde, werden nur Patienten mit einer Baseline und mindestens einer Post-Baseline-Messung in die Analyse einbezogen“ [21].

⁵⁾ GEE-Modell mit Veränderung der FPS-R von Baseline als abhängige Variable, Visite, Behandlungsarm, Interaktion zwischen Behandlungsgruppe und Studienvisite und Baseline-RSS als Faktoren und FPS-R bei Baseline als Kovariate.

Abkürzungen: FAS: Full Analysis Set; FPS-R: Faces Pain Scale - Revised; GEE: Generalized Estimation Equation; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; RSS: Rickets Severity Scale; SAP: Statistischer Analyseplan; SD: Standardabweichung; SE: Standardfehler.

3.4 Lebensqualität

In Studie 301 wurde Lebensqualität durch den SF-10 (Short Form Health Survey-10 for Children) erhoben. Aufgrund von Einschränkungen in den psychometrischen Eigenschaften werden die Ergebnisse der Fragebogenerhebung nicht dargestellt (s. Tabelle 7).

3.5 Sicherheit

Testpersonen aus der EU, Kanada, den USA und Australien konnten freiwillig an einer Extensionsperiode (ab Woche 66 bis maximal Woche 140) teilnehmen. In dieser Extensionsperiode erhielten alle Personen Burosumab. Zum Zeitpunkt des Datenschnitts hatten bereits 22 der Kinder der Kontrollgruppe in die unkontrollierte Extensionsperiode gewechselt und mindestens eine Dosis Burosumab erhalten. Da zu diesem Zeitpunkt die Daten unkontrolliert sind und die Expositionszeiten in der Extensionsperiode nur sehr kurz sind (ca. 1,5 Monate, vgl. Tabelle 15) werden in der Nutzenbewertung nur die Ergebnisse zur Sicherheit während der kontrollierten Behandlungsperiode dargestellt. In diesem Kapitel werden zudem ergänzend dentale Ereignisse dargestellt (s. Tabelle 6).

Mit Datenschnitt zu Woche 64 war die Behandlungsperiode für alle Testpersonen abgeschlossen. UE wurden ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung und bis zu 12 Wochen nach der letzten Gabe des Studienmedikaments gezählt. Ausgewertet wurden UE, die bei oder nach der ersten Dosis des Prüfprodukts auftraten und nicht bereits vor der ersten Dosisgabe bestanden (s. Tabelle 8).

Aus der zusammenfassenden Darstellung der UE (Tabelle 22) geht hervor, dass numerisch mehr UE im Interventions- als im Kontrollarm auftraten. Es trat im Interventionsarm ein Ereignis mehr mit Schweregrad ≥ 3 als im Kontrollarm auf. Der Unterschied war statistisch nicht signifikant. In beiden Studienarmen traten drei schwerwiegende unerwünschte Ereignisse (SUE) auf und es führte kein UE zum Abbruch der Studie oder der Studienmedikation.

Tabelle 22: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Behandlungsperiode bis Woche 64, Safety Set

Personen mit mindestens einem ...	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)	RR [95%-KI]; p-Wert
UE	29 (100)	27 (84,4)	– ¹⁾
UE CTCAE-Grad 3 und 4	4 (13,8)	3 (9,4)	1,5 [0,3; 7,6]; 0,6988 ²⁾
SUE	3 (10,3)	3 (9,4)	1,1 [0,2; 6,0]; 1,0000 ²⁾
UE, das zum Abbruch der Studie/ Studienmedikation führte	0	0	-

¹⁾ Patientenrelevanz nicht eindeutig beurteilbar.

²⁾ Angabe aus Modul 4 des Herstellerdossiers.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; KI: Konfidenzintervall; RR: Relatives Risiko; (S)UE: (Schwerwiegende/s) Unerwünschte/s Ereignis/se.

Häufige unerwünschte Ereignisse

In jeder Systemorganklasse traten numerisch mehr Ereignisse unter Behandlung mit Burosumab als unter der Kontrollbedingung auf (s. Tabelle 23). Am deutlichsten ist der Unterschied in den Systemorganklassen „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ und „Erkrankungen der Atemwege, des Brustraums und des Mediastinums“ (absoluter Unterschied > 40 %). Die häufigsten drei Preferred Terms in einem Studienarm waren Fieber (55,2 %), Husten (51,7 %) und Arthralgie (44,8 %), welche jeweils numerisch häufiger im Interventions- als im Kontrollarm auftraten. Effektschätzer wurden nicht vorgelegt.

Es ist anzumerken, dass einige der im Folgenden dargestellten Systemorganklassen und Preferred Terms XLH-Symptome umfassen und daher nicht auszuschließen ist, dass Symptome aus der Endpunktkategorie „Morbidität“ enthalten sind (s. Kapitel 4.6 Sicherheit).

Tabelle 23: UE mit Inzidenz bei > 10 % der Patientinnen und Patienten in einer der Behandlungsgruppen während der Behandlungsperiode bis Woche 64 der Studie 301, Safety Set

MedDRA-Systemorganklasse ¹⁾ Preferred Term	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	25 (86,2)	8 (25,0)
Fieber	16 (55,2)	6 (18,8)
Erythem an der Injektionsstelle	9 (31,0)	0
Reaktion an der Injektionsstelle	7 (24,1)	0
Pruritus an der Injektionsstelle	3 (10,3)	0
Schwellung an der Injektionsstelle	3 (10,3)	0
Ausschlag an der Injektionsstelle	3 (10,3)	0
Erkrankungen des Gastrointestinaltrakts	23 (79,3)	17 (53,1)
Erbrechen	12 (41,4)	8 (25,0)
Zahnkaries ²⁾	9 (31,0)	2 (6,3)
Diarrhö	7 (24,1)	2 (6,3)
Verstopfung	5 (17,2)	0
Übelkeit	3 (10,3)	1 (3,1)

MedDRA-Systemorganklasse¹⁾ Preferred Term	Burosumab N = 29 <i>n (%)</i>	Kontrolle N = 32 <i>n (%)</i>
Zahnschmerzen ²⁾	4 (13,8)	1 (3,1)
Schmerzen Oberbauch	3 (10,3)	3 (9,4)
Infektionen und parasitäre Erkrankungen	23 (79,3)	24 (75,0)
Nasopharyngitis	11 (37,9)	14 (43,8)
Zahnabszess ²⁾	8 (27,6)	3 (9,4)
Influenza	4 (13,8)	6 (18,8)
Infektion der oberen Atemwege	3 (10,3)	3 (9,4)
Erkrankungen der Atemwege, des Brustraums und des Mediastinums	21 (72,4)	9 (28,1)
Husten	15 (51,7)	6 (18,8)
Rhinorrhoea	7 (24,1)	2 (6,3)
Nasenkongestion	5 (17,2)	1 (3,1)
Oropharyngeale Schmerzen	5 (17,2)	1 (3,1)
Asthma	4 (13,8)	1 (3,1)
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen²⁾	17 (58,6)	15 (46,9)
Arthralgie	13 (44,8)	10 (31,3)
Schmerzen in den Extremitäten	11 (37,9)	10 (31,3)
Erkrankungen des Nervensystems	12 (41,4)	9 (28,1)
Kopfschmerzen	10 (34,5)	6 (18,8)
Erkrankungen der Haut und des Unterhautzellgewebes	11 (37,9)	4 (12,5)
Ausschlag	3 (10,3)	2 (6,3)
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	10 (34,5)	2 (6,3)
Kontusion	4 (13,8)	0
Sturz	3 (10,3)	0
Erkrankungen des Ohrs und des Labyrinths	6 (20,7)	3 (9,4)
Ohrenschmerzen	4 (13,8)	1 (3,1)
Untersuchungen	8 (27,6)	4 (12,5)
Vitamin D vermindert	6 (20,7)	1 (3,1)
Stoffwechsel- und Ernährungsstörungen	6 (20,7)	3 (9,4)
Vitamin-D-Mangel	5 (17,2)	1 (3,1)
Erkrankungen des Immunsystems	5 (17,2)	5 (15,6)
Saisonale Allergie	4 (13,8)	2 (6,3)
Kongenitale, familiäre und genetische Erkrankungen²⁾	3 (10,3)	1 (3,1)

¹⁾ MedDRA-Version 18.1.

²⁾ Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidität“.

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; UE: Unerwünschte/s Ereignis/se.

Dentale Ereignisse

In Studie 301 wurden dentale Ereignisse vom klinischen Studienpersonal entweder als Ereignis der Morbidität oder der Sicherheit eingestuft (s. Tabelle 6). Dentale Ereignisse, die der Morbidität zugeordnet wurden, werden in dieser Nutzenbewertung ergänzend in der Kategorie Sicherheit dargestellt.

Zu Woche 64 zeigte sich gegenüber Baseline eine numerische Steigerung jeglicher Zahnbeschwerden für beide Studienarme: im Interventionsarm von 2 auf 12 Ereignisse und im Kontrollarm von 1 auf 5. Im Interventionsarm traten zu Woche 64 mit einem Unterschied von 7 Ereignissen numerisch mehr Zahnbeschwerden auf als in der Kontrollgruppe (Tabelle 24).

Tabelle 24: Dentale Ereignisse¹⁾ während der Behandlungsperiode in Studie 301, Safety Set

UX023-CL301	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
<i>Jegliche Zahnbeschwerden zu Baseline</i>	2 (6,9)	1 (3,1)
Zahnabszess	1 (3,4)	1 (3,1)
Zahnextraktion (durch Abszess oder Karies)	1 (3,4)	0
Wurzelbehandlung	0	0
Zahnkaries	1 (3,4)	0
Gingivitis	0	0
<i>Jegliche Zahnbeschwerden zu Woche 64</i>	12 (41,4)	5 (15,6)
Zahnabszess	10 (34,5)	2 (6,3)
Zahnextraktion (durch Abszess oder Karies)	5 (17,2)	3 (9,4)
Wurzelbehandlung	3 (10,3)	0
Zahnkaries	6 (20,7)	1 (3,1)
Gingivitis	0	0

¹⁾ Die Benennung relevanter Ereignisse unterscheidet sich in den Studienunterlagen (s. Tabelle 6).

Unerwünschte Ereignisse mit CTCAE-Grad ≥ 3

In der Behandlungsperiode der Studie 301 traten insgesamt 4 UE mit CTCAE-Grad 3 im Interventionsarm und 3 Ereignisse im Kontrollarm auf (s. Tabelle 25). Kein UE wurde dem Grad 4 bzw. 5 zugeordnet.

Tabelle 25: UE mit CTCAE-Grad 3 während der Behandlungsperiode in Studie 301, Safety Set

MedDRA-Systemorganklasse¹⁾, Preferred Term	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Infektionen und parasitäre Erkrankungen , virale Gastroenteritis	1 (3,4)	0
Untersuchungen , Ketonkörper im Urin	1 (3,4)	0
Erkrankungen der Muskel, des Skeletts und des Bindegewebes²⁾ , Arthralgie	1 (3,4)	1 (3,1)
Nieren- und Harnwegserkrankungen , Dysurie	1 (3,4)	0
Angeborene, familiäre und genetische Störungen²⁾ , Kraniosynostose	0	1 (3,1)
Störungen des Immunsystems , Nahrungsmittelallergie	0	1 (3,1)

²⁾ MedDRA-Version 18.1.

³⁾ Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidität“.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; MedDRA: Medical Dictionary for Regulatory Activities; UE: Unerwünschte/s Ereignis/se.

Schwerwiegende unerwünschte Ereignisse

Sowohl im Interventions- als auch im Kontrollarm erlebten jeweils 3 Personen je ein SUE während der Behandlungsperiode; kein SUE trat mehrfach auf (s. Tabelle 26).

Tabelle 26: SUE während der Behandlungsperiode in Studie 301, Safety Set

MedDRA-Systemorganklasse ¹⁾ , Preferred Term	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Infektionen und parasitäre Erkrankungen, Virale Infektion	1 (3,4)	0
Erkrankungen des Nervensystems, Migräne	1 (3,4)	0
Angeborene, familiäre und genetische Störungen ²⁾ , Kraniosynostose	1 (3,4)	1 (3,1)
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen ²⁾ , Verformungen der Knie	0	1 (3,1)
Nieren- und Harnwegserkrankungen, Hämaturie	0	1 (3,1)

⁴⁾ MedDRA-Version 18.1.

⁵⁾ Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidity“.

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; SUE: Schwerwiegende/s unerwünschte/s Ereignis/se.

Unerwünschte Ereignisse von besonderem Interesse

UE von besonderem Interesse waren im SAP a priori festgelegt. Zwischen Studienbericht und SAP besteht die Abweichung, dass laut Studienbericht „Reaktionen an der Injektionsstelle“ kein Ereignis von Interesse im Kontrollarm waren. Diese Einschränkung findet sich jedoch nicht im SAP, in welchem UE von besonderem Interesse erstmals definiert wurden. Insgesamt traten bei 15 Personen im Interventionsarm Ereignisse in dieser Kategorie auf. Bei 11 Personen aus dem Interventions- und 6 Personen aus dem Kontrollarm traten zudem Hypersensibilitäten auf (s. Tabelle 27).

Tabelle 27: UE von besonderem Interesse während der Behandlungsperiode in Studie 301, Safety Set

Personen mit UE von besonderem Interesse ¹⁾	Burosumab N = 29 n (%)	Kontrolle N = 32 n (%)
Reaktionen an der Injektionsstelle	15 (51,7)	0 ²⁾
Hypersensibilität	11 (37,9)	6 (18,8)
Ektopische Mineralisierung	0	0
Hyperphosphatämie	0	0
Restless-Legs-Syndrom	0	0

¹⁾ Abweichungen zu Tabelle 23 ergeben sich aus der Erhebung der UE von besonderem Interesse über HLT, SMQ und durch vom pU erstellte Kategorisierungen von Preferred Terms gemäß MedDRA (Version 18.1.) (s. Tabelle 8).

²⁾ Laut Studienbericht waren Reaktionen an der Injektionsstelle kein UE von besonderem Interesse in der Kontrollgruppe. Diese Einschränkung findet sich nicht in SAP oder Studienprotokoll. In den statistischen Outputs werden 0 Ereignisse in der Kontrollgruppe aufgeführt.

Abkürzungen: HLT: High Level Term; MedDRA: Medical Dictionary for Regulatory Activities; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; SMQ: Standardised MedDRA Query; UE: Unerwünschte/s Ereignis/se.

3.6 Subgruppen

Der pU legt mit seinem Dossier post hoc durchgeführte Interaktionstests vor. Gemäß der unter 2.4 genannten Kriterien zeigten sich für keine der untersuchten Subgruppen statistisch signifikante Interaktionen mit den Subgruppenmerkmalen.

4 Diskussion der Methodik und Ergebnisse

4.1 Zulassungsstatus und Zulassungspopulation von Burosumab

Gemäß Fachinformation ist Burosumab indiziert zur Behandlung von Kindern ab 1 Jahr und Jugendlichen in der Skelettwachstumsphase mit X-chromosomaler Hypophosphatämie (XLH) und röntgenologischem Nachweis einer Knochenerkrankung [20]. Die Therapie sieht die subkutane Injektion des Antikörpers ca. alle 2 Wochen durch eine Ärztin, einen Arzt oder eine andere geschulte medizinische Fachkraft vor. Die empfohlene Anfangsdosis wird mit 0,4 mg/kg Körpergewicht, die Erhaltungstherapie mit 0,8 mg/kg angegeben. Eine zeitgleiche, konservative Behandlung mit Phosphat und aktivem Vitamin D ist kontraindiziert [20].

Burosumab wurde 2018 durch die EMA unter „Besonderen Bedingungen“ zugelassen. Als Bedingung der Zulassung wurde u. a. die Vorlage der finalen Ergebnisse aus den Phase-II-Studien UX023-CL201 und UX023-CL205 sowie von zum Zeitpunkt der Zulassung noch nicht verfügbaren Daten der Studie 301 zum Nachweis der Wirksamkeit auferlegt [6]. Studie 301 bildet die Grundlage der vorliegenden Nutzenbewertung, welche im Zuge des bis 31. Oktober 2019 befristeten Beschlusses zum erstmaligen Inverkehrbringen von Burosumab vom 4. Oktober 2018 [10] erstellt wurde.

Die vorliegende Studie 301 rekrutierte Kinder im Alter von 1 bis 12 Jahren. Für die Zulassungspopulation der 13- bis 17-Jährigen liegen keine Daten vor.

An Studie 301 nahmen keine Kinder aus Deutschland teil. Die Ergebnisse der Testpersonen aus den beteiligten Ländern erscheinen jedoch auf den deutschen Versorgungskontext übertragbar.

4.2 Design und Methodik der Studie

Bei Studie 301 handelt es sich um eine offene, randomisiert kontrollierte, multizentrische Studie der Phase III im Parallelgruppendesign (1:1) mit zwei Behandlungsarmen. In dieser Studie wird die Wirksamkeit und Sicherheit von Burosumab gegenüber einer Therapie mit oralem Phosphat und aktivem Vitamin D bei XLH-erkrankten Kindern im Alter von 1 bis 12 Jahren untersucht. Als primärer Endpunkt der Studie wurde die Veränderung im RGI-C-Gesamtwert von Baseline zu Woche 40 festgelegt.

Die Studie gliederte sich in eine Screening-, eine Behandlungs- und eine Extensionsperiode. Eingereicht wurden Studienunterlagen mit Datenschnitt zu Studienwoche 64. Zu diesem Zeitpunkt hatten alle Studienteilnehmenden die Behandlungsperiode abgeschlossen. Für Studienteilnehmende aus Europa, den USA, Kanada und Australien schließt sich nach Woche 64 eine unkontrollierte Extensionsperiode (bis Woche 140) an, in der alle Teilnehmenden Burosumab erhalten. In Japan und Südkorea bestand diese Möglichkeit nicht; hier endete die Studie mit Ende der Behandlungsperiode zu Woche 64.

Studie 301 rekrutierte an XLH erkrankte Mädchen und Jungen mit radiologischem Nachweis einer Rachitis, definiert als RSS-Gesamtscore von $\geq 2,0$, nachgewiesener PHEX-Mutation und biochemischen Befunden, die mit XLH assoziiert sind. Zu Baseline zeigten sich zwischen den Studienarmen im Wesentlichen gleiche Verteilungen, u. a. bei Alter, Geschlecht, Körpergröße und RSS-Gesamtwert. Die Studienpopulation ist durch einen nahezu ausgeglichenen Anteil an Mädchen (55 % im Interventionsarm, 56 % im Kontrollarm) und Jungen (45 % im Interventionsarm, 44 % im Kontrollarm) gekennzeichnet. Durchschnittlich waren die Kinder zu Baseline im Interventionsarm

5,8 Jahre alt und im Kontrollarm 6,3 Jahre alt. Hinsichtlich der alters- und geschlechtsadjustierten Perzentile der Stehhöhe zu Baseline zeigt sich eine Diskrepanz zwischen Mittelwert und Median in beiden Studienarmen, die in Anbetracht des geringen Stichprobenumfangs jedoch nicht ungewöhnlich erscheint. Die Angabe der Stehhöhe als z-Score zeigt diesen Unterschied durch die Transformation der Werte nicht. Die Darstellung alters- und geschlechtsadjustierter Maßzahlen wird grundsätzlich begrüßt.

Eine Therapie mit Burosumab wird laut Fachinformation [20] mit einer empfohlenen Anfangsdosis von 0,4 mg/kg Körpergewicht begonnen. Die übliche Erhaltungsdosis liegt bei 0,8 mg/kg alle zwei Wochen mit einer Höchstdosis von 90 mg. Zur Therapiesteuerung dient der altersbezogene Referenzbereich des Nüchtern-Serumphosphats, der als Ziel der Therapie erreicht werden soll. In Studie 301 begann die Therapie mit Burosumab mit einer Anfangsdosis von 0,8 mg/kg Körpergewicht alle zwei Wochen und konnte bei Bedarf auf 1,2 mg/kg Körpergewicht gesteigert werden. Während der Behandlungsperiode verblieben 72 % der Kinder im Burosumab-Arm bei dieser Anfangsdosis, bei 28 % war eine Dosissteigerung notwendig. Der Beginn der Therapie mit 0,8 mg/kg entspricht damit nicht der Fachinformation. Die Wahl der Startdosis wird mit Erfahrungen aus der Dosisfindungsstudie UX023-CL201 der Phase II begründet. Auch in dieser Studie war als Startdosis 0,8 mg/kg Körpergewicht festgesetzt. Daraus resultierende Konsequenzen hinsichtlich der Übertragbarkeit auf die Therapiesituation in Deutschland sind gerade mit Blick auf die individuelle Titration zur Erreichung des altersbezogenen Referenzbereichs für Serumphosphat nicht abschätzbar.

Während der Behandlungsphase berichteten Studienteilnehmende, die mit Burosumab behandelt wurden, häufiger die Einnahme von Begleitmedikamenten als Studienteilnehmende im Kontrollarm. Darunter fielen neben Paracetamol und Ibuprofen auch Husten- und Erkältungsmittel und systemische Antibiotika. Insbesondere die Einnahme der Schmerzmittel könnte zu einer Verzerrung der Ergebnisse zur Wirksamkeit und Sicherheit von Burosumab geführt haben. Es ist unklar, welche Gründe für die Einnahme der Begleitmedikation in beiden Studienarmen vorlagen und ob sich diese Gründe zwischen den Studienarmen unterschieden.

Das Studienprotokoll der Studie 301 wurde einmal, ca. 10 Monate nach Studienbeginn, geändert. Relevante Änderungen umfassten u. a. Änderungen an den Einschlusskriterien und die Änderung der eingesetzten Fragebogenversion des PROMIS. Aus dem Studienplan ergibt sich, dass die Rekrutierung zum Zeitpunkt der Protokolländerung jedoch bereits abgeschlossen war. Auch laut Herstellermodul wurde der bereits stattfindende Einsatz der neueren PROMIS-Version mit der Protokolländerung dokumentiert. Damit ergibt sich, dass für einen unbestimmten Zeitraum die Studiendurchführung nicht nach Studienprotokoll stattfand. Dieses Vorgehen erscheint ungewöhnlich für eine klinische Studie der Phase III und erhöht insgesamt die Unsicherheit in der Einschätzung der Ergebnisse.

Mit dem Protokollamendment wurde zudem die Nachbeobachtung der Sicherheitsendpunkte geändert: Die auf 12 Wochen nach der letzten Gabe des Studienmedikaments angesetzte Visite konnte für Studienteilnehmende, die zu Woche 64 aus der Studie austraten, nun auch außerhalb von Studie 301 erfolgen. Es ist unklar, wie viele Kinder außerhalb von Studie 301 nachbeobachtet wurden. Insgesamt traten 86 % der Kinder aus dem Burosumab-Arm und 81 % aus dem Kontrollarm in die Extensionsperiode ein und benötigten dadurch keine Nachbeobachtungsvisite. Für max. 16 % der Studienpopulation fehlen damit jedoch Informationen nach Behandlungsende, was die Vertrauenswürdigkeit der Ergebnisse zur Sicherheit nach Woche 64 einschränken würde.

Der pU hat zusätzlich ein COA-Dossier [30] eingereicht, welches im klinischen Entwicklungsprogramm zu Burosumab die Erhebung von Endpunkten übergreifend beschreibt. Zudem sind eigene Untersuchungen zu den psychometrischen Eigenschaften der Instrumente im Erkrankungsbild dargestellt. Der ausführliche, konzeptuelle Umriss und die soliden Untersuchungen der Instrumente werden ausdrücklich begrüßt. Sie ermöglichen eine abschließende Einschätzung zur Geeignetheit der Instrumente, die andernfalls nicht möglich gewesen wäre. Gleichzeitig sei angemerkt, dass der Stellenwert der COA-Dossiers gegenüber dem Studienprotokoll unklar ist, insbesondere bei abweichenden Informationen zum Einsatz der Instrumente in Studie 301. Zudem können Änderungen zwischen verschiedenen Versionen des Dokuments nur unzureichend nachvollzogen werden. Die Aussagekraft der psychometrischen Untersuchungen ist dadurch jedoch nicht eingeschränkt.

Die statistische Analyse wurde im SAP geplant und ist nachvollziehbar beschrieben. Es ist nicht berichtet, inwiefern die statistischen Modellannahmen erfüllt waren. Die vorgelegten Subgruppenanalysen (u. a. hinsichtlich Alter, Geschlecht) lassen basierend auf der verfügbaren Datengrundlage keine belastbaren Aussagen zu.

Die Vertrauenswürdigkeit der Studienergebnisse ist insgesamt aufgrund des offenen Studiendesigns limitiert und erscheint besonders hinsichtlich der subjektiven Endpunkte gering.

4.3 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserfassung dokumentiert. Im Verlauf der Studie 301 ist keine Person verstorben.

4.4 Morbidität

Zur Bewertung der Wirksamkeit in der Kategorie Morbidität wurden die Endpunkte Körpergröße, Gehfähigkeit mittels 6MWT, Beeinträchtigung durch Schmerzen, körperliche Funktion und Fatigue mittels PROMIS sowie Schmerzintensität mittels FPS-R herangezogen.

Körpergröße

Im z-Score der Stehhöhe/Liegelänge zeigte sich während der 64-wöchigen Behandlungsperiode eine mittlere Verbesserung um 0,2 im Interventionsarm und keine Veränderung im Kontrollarm. Die LS-Mean-Differenz der Veränderungen beträgt 0,1 [95%-KI: [0,0; 0,3]] zugunsten der Therapie mit Burosumab. Diese Differenz war statistisch signifikant mit $p = 0,049$. Im Interventionsarm fehlte von einem Kind die Messung zu Baseline, sodass keine Veränderung für dieses Kind berechnet wurde. Es finden sich keine Angaben, warum diese Messung fehlte; da sie jedoch bereits zu Baseline fehlte ist nicht von einer systematischen Verzerrung auszugehen. Sensitivitätsanalysen wurden nicht vorgelegt. Vor dem Hintergrund der geringen Ausprägung in der Differenz zwischen den Studienarmen zeigte sich auch gemäß Hedges' g keine klinische Relevanz in der Differenz der Veränderungen.

Die Veränderung in den alters- und geschlechtsadjustierten Perzentilen der Stehhöhe/Liegelänge und der absoluten Sitzhöhe wurde deskriptiv ausgewertet. Numerisch verbesserten sich die Kinder im Interventionsarm im Median um 0,2 Wachstumsperzentile, im Kontrollarm zeigte sich hingegen im Median keine Veränderung der Wachstumsperzentile. Unterschiede zwischen den Studienarmen können aufgrund der deskriptiven Auswertung nicht bewertet werden. In der absoluten Sitzhöhe zeigte sich im Interventionsarm eine Veränderung im Median um 3,5 cm, im Kontrollarm um

3,8 cm. Für jeweils 3 Kinder je Studienarme fehlten die Informationen zur Sitzhöhe. Diese Veränderungen von Baseline zu Woche 64 innerhalb eines Studienarms sind durch das zunehmende Alter der Kinder nicht interpretierbar.

Gehfähigkeit (6MWT)

Es zeigte sich eine statistisch signifikant größere Veränderung um 43 m in der Laufstrecke zugunsten von Burosumab gegenüber der Kontrollgruppe. Es existiert keine validierte klinische Relevanzschwelle im Erkrankungsbild. Gemäß Hedges' g ist die gezeigte Differenz nicht klinisch relevant. Im prozentualen Anteil der erwarteten Laufstrecke zeigte sich kein statistisch signifikanter Unterschied zwischen den Studienarmen, sodass der gezeigte Vorteil in der absoluten Laufstrecke auch vor dem Hintergrund der absolvierten Laufstrecke zu Baseline kritisch zu betrachten ist.

Die Analysen wurden nicht – wie a priori geplant – für Alter zu Baseline adjustiert. Zudem bestehen Unterschiede in der 6MWT-Distanz zu Baseline zwischen den Studienarmen (366 m im Burosumab-, 451 m im Kontrollarm). Zwar wurde diese Variable als Kovariate im Modell berücksichtigt, inwiefern der Unterschied durch diese Adjustierung jedoch ausgeglichen werden kann ist unklar.

Unsicherheiten ergeben sich zudem daraus, dass nicht bekannt ist, wie viele Personen in der Analyse berücksichtigt wurden. Es fehlten für 2 Kinder des Burosumab-Arms Daten: Ein Kind absolvierte den 6MWT nicht zu Baseline, ein Kind nicht zu Woche 40. Im Herstellermodul gibt der pU an, dass bei der Auswertung von Veränderungen nur Kinder mit einer Baseline- und mindestens einer Post-Baseline-Messung in der Analyse berücksichtigt werden. Damit wären 14 der 15 Kinder im Burosumab-Arm in der Auswertung. In den statistischen Outputs ist die Analysepopulation jedoch mit 15 Personen angegeben, sodass unklar ist, wie viele Personen tatsächlich in der Analyse berücksichtigt wurden.

Beeinträchtigung durch Schmerzen, körperliche Funktion und Fatigue: PROMIS

Bei PROMIS bestehen Einschränkungen in der Nachvollziehbarkeit der Operationalisierung. Zudem stellt der pU die Ergebnisse der Versionen aus Selbst- und Fremdeinschätzung gemeinsam dar. Es finden sich in der PROMIS-Methodik keine Informationen dazu, ob dies adäquat ist. Der Anteil an Selbst- und Fremdeinschätzungen ist jedoch in den Studienarmen gleich, sodass die Ergebnisse berücksichtigt werden konnten.

In den 3 Domänen des PROMIS zur Erhebung von Beeinträchtigung durch Schmerzen, körperlicher Funktion und Fatigue zeigten sich zwischen den Studienarmen keine statistisch signifikanten Unterschiede in den Veränderungen von Baseline bis Woche 64. Responderanalysen wurden nicht vorgelegt.

Schmerzintensität (FPS-R)

Die Erhebung der Schmerzintensität zeigte zu Baseline in beiden Studienarmen nur sehr geringe Ausprägungen. In beiden Studienarmen wurden keine Veränderungen im Verlauf der Behandlungsperiode bis Woche 64 beobachtet und demzufolge kein statistisch signifikanter Unterschied zwischen den Studienarmen festgestellt. Da das Instruments aktuell bestehende Schmerzen erfragt, ist es wahrscheinlich, dass das chronische Schmerzbild der XLH nicht vollumfänglich abgebildet ist.

Zusammenfassende Einschätzung zur Morbidität

Es zeigten sich statistisch signifikante Vorteile für Burosumab gegenüber der Kontrollgruppe in der Veränderung im z-score der Körperhöhe/Liegelänge und in der absoluten Laufstrecke des 6MWT. Die klinische Relevanz dieser Effekte ist unklar.

Es zeigten sich keine Unterschiede zwischen den Studienarmen im prozentualen Anteil der erwarteten Laufstrecke, der Beeinträchtigung durch Schmerzen, der körperlichen Funktion und Fatigue mittels PROMIS und der Schmerzintensität mittels FPS-R.

Die Ergebnissicherheit ist aufgrund der geringen Studiengröße und bei den subjektiven Endpunkten 6MWT, PROMIS und FPS-R aufgrund teilweise fehlender Informationen zur Auswertung, des höheren Anteils an Begleitmedikation im Interventionsarm und des offenen Studiendesigns limitiert. Mögliche Verzerrungen scheinen sich in der Gesamtschau bei den subjektiven Endpunkten zugunsten von Burosumab auszuwirken.

4.5 Lebensqualität

Aufgrund von Einschränkungen in den psychometrischen Eigenschaften sind die Ergebnisse der Erhebung des SF-10 in der Nutzenbewertung nicht berücksichtigt (s Tabelle 7). In der Kategorie Lebensqualität liegen damit keine validen Ergebnisse vor.

4.6 Sicherheit

In der Nutzenbewertung werden die Auswertungen der Sicherheit während der Behandlungsperiode bis Woche 64 berücksichtigt. Danach schließt sich die unkontrollierte Extensionsperiode an, in der alle Studienteilnehmenden Burosumab erhalten. Da ab diesem Zeitpunkt keine vergleichenden Daten mehr vorliegen und die zusätzlichen Expositionszeiten in der Extensionsperiode nur sehr kurz waren, werden diese Daten in der Nutzenbewertung nicht berücksichtigt.

Das Auftreten unerwünschter Ereignisse wurde ab Unterzeichnung der Einverständniserklärung erhoben. Ausgewertet wurden hingegen erst Ereignisse nach Gabe der ersten Dosis der Studienmedikation. Da es sich bei XLH um eine langsam progrediente Erkrankung handelt und angenommen wird, dass die Zeiträume zwischen Unterzeichnung der Einverständniserklärung und Gabe der ersten Dosis in beiden Studienarmen ähnlich lang waren, wird der Endpunkt dennoch bis Woche 64 als valide bewertet. Konkrete Angaben zu den Zeiträumen fehlen jedoch.

Die Durchführung der Visiten unterscheidet sich zwischen den Studienarmen: Im Interventionsarm konnten außer in Japan und Südkorea je nach Nähe und Patientenpräferenz Hausvisiten durchgeführt werden; im Kontrollarm sollten stattdessen telefonische Abfragen stattfinden. Es finden sich keine Angaben, wie viele Personen im Interventionsarm wie viele Hausvisiten in Anspruch nahmen und ob telefonische Abfragen in der gleichen Häufigkeit stattfanden. Es ist auch nicht beschrieben, ob in Japan und Südkorea für alle Teilnehmenden telefonische Abfragen vorgesehen waren, da für diese Personen keine Option für Hausvisiten bestand. Je nach Konstellation ist eine stärkere Erfassung entweder im Interventions- oder im Kontrollarm denkbar. Grundsätzlich erscheint auch das unterschiedliche Setting der Erhebung bedenkenswert, wobei dies als nachteilig für den Interventionsarm mit Burosumab beurteilt wird.

Innerhalb der Behandlungsperiode bis Woche 64 trat bei allen Personen mindestens ein UE auf. Sowohl im Hinblick auf UE mit CTCAE-Grad 3 und 4 und SUE zeigte sich kein statistisch

signifikanter Unterschied zwischen den Studienarmen. Kein UE führte zum Abbruch der Studie oder der Studienmedikation.

Häufige UE ($\geq 10\%$) jedweden Schweregrades traten besonders in den Systemorganklassen „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“, „Erkrankungen des Gastrointestinaltrakts“ und „Infektionen und parasitäre Erkrankungen“ auf. In allen Systemorganklassen traten numerisch mehr UE unter Burosumab als unter Kontrolle auf. Effektschätzer für UE auf Ebene der Systemorganklassen wurden nicht vorgelegt.

Es wurden UE der Systemorganklasse „Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen“ sowie der Preferred Terms „Zahnschmerzen“, „Zahnkaries“ und „Zahnabszess“ berichtet. Ereignisse dieser Kategorien sind jedoch auch typische Symptome der hier betrachteten Grunderkrankung, welche der Endpunktkategorie „Morbidität“ zuzuordnen wären. Es ist allerdings nicht möglich, einzelne Symptome wie Schmerzen in den Extremitäten aus den betreffenden Systemorganklassen zu separieren. Die vom pU durchgeführte Kategorisierung von dentalen Ereignissen in Morbidität und Sicherheit weist ebenfalls deutliche Mängel in der Operationalisierung auf. In der ergänzenden Darstellung dieses Endpunkts in der Kategorie Sicherheit traten numerisch mehr Ereignisse zu Woche 64 im Interventions- als im Kontrollarm auf.

Vom pU wurden a priori Reaktionen an der Injektionsstelle, Hypersensibilität, ektopische Mineralisierung, Hyperphosphatämie und Restless-Legs-Syndrom als UE von besonderem Interesse definiert. Die Auswahl wurde nicht begründet. Infusionsbedingte Reaktionen sind bei 52 % der Kinder im Interventionsarm aufgetreten. Laut Studienbericht war dies kein UE von besonderem Interesse im Kontrollarm. Auch wenn dies aus Mangel an Injektionen im Kontrollarm plausibel erscheint, ist diese Einschränkung nicht a priori dokumentiert. Hypersensibilitäten wiesen 38 % der Kinder im Interventionsarm und 19 % im Kontrollarm auf. Weitere UE von besonderem Interesse wurden nicht beobachtet.

Zusammenfassende Einschätzung zur Sicherheit

Bis Woche 64 erscheinen folgende Faktoren die Vertrauenswürdigkeit der Ergebnisse wesentlich zu beeinträchtigen: Die unbekannte Häufigkeit der Hausvisiten im Gegensatz zu telefonischen Abfragen und der erhöhte Anteil an Begleitmedikation im Interventionsarm mit jeweils unklarer Richtung eines Vor- oder Nachteils sowie das unterschiedliche Setting der Visiten zum Nachteil von Burosumab. Weitere Einschränkungen ergeben sich aus dem limitierten Stichprobenumfang.

Zudem ist unklar, inwiefern einige der berichteten UE auch Symptome der Grunderkrankung widerspiegeln.

Vor diesem Hintergrund traten bei gleichen Beobachtungszeiten insgesamt mehr unerwünschte Ereignisse im Interventionsarm im Vergleich zum Kontrollarm auf. Eine abschließende Bewertung des Sicherheitsprofils von Burosumab ist aufgrund der geringen Stichprobe besonders in Hinblick auf schwere und schwerwiegende unerwünschte Ereignisse nicht möglich.

5 Anforderung an eine qualitätsgesicherte Anwendung

Die Vorgaben der Fachinformation sind zu berücksichtigen. Die Behandlung mit Burosumab muss von einem Arzt, einer Ärztin eingeleitet werden, der/die Erfahrung in der Therapie von Patientinnen und Patienten mit Knochenstoffwechselerkrankungen hat.

6 Zusammenfassung der Nutzenbewertung

Burosumab ist zugelassen für die Behandlung von Kindern ab 1 Jahr und Jugendlichen in der Skelettwachstumsphase mit XLH mit röntgenologischen Nachweis einer Knochenerkrankung. Die Nutzenbewertung von Burosumab basiert auf Studie 301, welche Teil der Auflagen für die Zulassung unter Bedingungen war. Es handelt sich um eine randomisierte, aktiv kontrollierte Studie der Phase III.

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

Tabelle 28: Zusammenfassende Darstellung der Ergebnisse der Studie UX023-CL301 für die Behandlungsperiode

Endpunkt	Burosumab (N = 29) vs. Kontrolle (N = 32)	Effekt
Morbidität		
Veränderung der Stehhöhe/Liegelänge zwischen Baseline und Woche 64 (z-Score) ¹⁾ LS-Mean-Differenz [95%-KI]; p-Wert ²⁾³⁾ Hedges' g [95%-KI]	0,1 [0,0; 0,3]; 0,049 0,5 [-0,03; 1,0] ⁴⁾	$\uparrow\uparrow$
Veränderung der 6MWT-Distanz zwischen Baseline und Woche 64 (Meter) ⁵⁾ LS-Mean-Differenz [95% KI]; p-Wert ⁶⁾⁷⁾ Hedges' g [95%-KI]	43,2 [2,3; 84,1]; 0,0383 0,8 [0,1; 1,5] ⁴⁾	\uparrow
Veränderung des Anteils der erwarteten 6MWT-Distanz zwischen Baseline und Woche 64 (%) ⁵⁾ LS-Mean-Differenz [95%-KI]; p-Wert ⁶⁾⁷⁾	6,2 [-0,7; 13,2]; 0,0781	\leftrightarrow
Veränderung der Symptome mittels PROMIS zwischen Baseline und Woche 64 ⁵⁾ LS-Mean-Differenz [95%-KI]; p-Wert ⁸⁾	Paediatric Pain Interference Score ⁹⁾ : -2,3 [-6,6; 2,1]; 0,3091 Physical Function Mobility Score ¹⁰⁾ : 1,9 [-1,8; 5,6]; 0,3145 Fatigue Score ¹¹⁾ : -1,1 [-6,2; 4,1]; 0,6810	\leftrightarrow
Veränderung des FPS-R-Werts für Schmerzintensität ⁴⁾⁵⁾¹²⁾ LS-Mean-Differenz [95%-KI]; p-Wert ¹³⁾	0,1 [-0,6; 0,7]; 0,8786	\leftrightarrow
Sicherheit		
UE CTCAE-Grad 3 und 4 Burosumab vs. Kontrollarm, n (%) RR [95%-KI]; p-Wert	4 (13,8) vs. 3 (9,4) 1,5 [0,3; 7,6]; 0,6988 ⁴⁾	\leftrightarrow
SUE Burosumab vs. Kontrollarm, n (%) RR [95%-KI]; p-Wert	3 (10,3) vs. 3 (9,4) 1,1 [0,2; 6,0]; 1,0000 ⁴⁾	\leftrightarrow
UE, das zum Abbruch der Studienmedikation führte Burosumab vs. Kontrollarm, n (%)	Keine Ereignisse aufgetreten.	n. b.
UE, das zum Tod führte Burosumab vs. Kontrollarm, n (%)	Keine Ereignisse aufgetreten.	n. b.

Endpunkt	Burosumab (N = 29) vs. Kontrolle (N = 32)	Effekt
UE von besonderem Interesse, n (%)		
<i>Reaktionen an der Injektionsstelle</i> Burosumab vs. Kontrollarm	15 (51,7) vs. 0 ¹⁴⁾	n. b.
<i>Hypersensibilität</i> Burosumab vs. Kontrollarm	11 (37,9) vs. 6 (18,8)	n. b.
<i>Ektopische Mineralisierung</i> Burosumab vs. Kontrollarm	Keine Ereignisse aufgetreten.	n. b.
<i>Hyperphosphatämie</i> Burosumab vs. Kontrollarm	Keine Ereignisse aufgetreten.	n. b.
<i>Restless-Legs-Syndrom</i> Burosumab vs. Kontrollarm	Keine Ereignisse aufgetreten.	n. b.

- ¹⁾ Auswertung in der FAS-Population (Burosumab: N = 29, Kontrolle: N = 32), welche dem ITT-Prinzip entspricht.
- ²⁾ GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und Baseline-Alter und Baseline-Stehhöhe/-Liegelänge (z-Score) als kontinuierliche Kovariaten.
- ³⁾ Für ein Kind im Burosumab-Arm fehlte die Angabe zu Baseline. Es wurde nicht in der Analyse berücksichtigt.
- ⁴⁾ Angabe aus dem Herstellermodul entnommen, da nicht verfügbar im Studienbericht.
- ⁵⁾ Auswertung im Full Analysis-Subset mit allen Personen ≥ 5 Jahre zu Baseline. Burosumab: N = 15, Kontrolle: N = 20.
- ⁶⁾ GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und 6MWT zu Baseline als kontinuierliche Kovariate. Laut SAP war auch eine Adjustierung für Baseline-Alter vorgesehen, was jedoch nicht umgesetzt wurde.
- ⁷⁾ Es fehlten für 2 Kinder des Burosumab-Arms Daten. Es ist unklar, wie viele Kinder in der Auswertung berücksichtigt wurden.
- ⁸⁾ GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Stratifizierungsfaktor Baseline-RSS als unabhängige Variablen und Baseline-PROMIS-Domänen-Score als kontinuierliche Kovariate.
- ⁹⁾ Höhere Werte sind gleichbedeutend mit einer stärkeren Beeinträchtigung durch Schmerzen.
- ¹⁰⁾ Höhere Werte sind gleichbedeutend mit einer besseren körperlichen Aktivität.
- ¹¹⁾ Höhere Werte sind gleichbedeutend mit einer stärkeren Fatigue.
- ¹²⁾ Höhere Werte sind gleichbedeutend mit einer stärkeren Schmerzintensität, Skala reicht von 0 bis 10.
- ¹³⁾ GEE-Modell mit Veränderung von Baseline als abhängige Variable, Studienvisite, Behandlungsarm, Interaktion zwischen Behandlungsgruppe und Studienvisite und Baseline-RSS als Faktoren und FPS-R bei Baseline als Kovariate, mit austauschbarer Kovarianzstruktur.
- ¹⁴⁾ Laut Studienbericht waren Reaktionen an der Injektionsstelle kein UE von besonderem Interesse in der Kontrollgruppe. Diese Einschränkung findet sich nicht im SAP oder Studienprotokoll.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; CTCAE: Common Terminology Criteria for Adverse Events; FAS: Full Analysis Set; FPS-R: Faces Pain Scale - Revised; GEE: Generalized Estimation Equation; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; n. b.: nicht bewertbar; PROMIS: Patient-Reported Outcomes Measurement Information System; RR: Relatives Risiko; RSS: Rickets Severity Scale; SAP: Statistischer Analyseplan; (S)UE: (Schwerwiegende/s) Unerwünschte/s Ereignis/se.

Referenzen

1. **Bieri D, Reeve RA, Champion GD, Addicoat L, Ziegler JB.** The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. *Pain* 1990;41(2):139-150.
2. **Bohannon RW, Crouch R.** Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. *J Eval Clin Pract* 2017;23(2):377-381.
3. **Carpenter TO, Imel EA, Holm IA, Jan de Beur SM, Insogna KL.** A clinician's guide to X-linked hypophosphatemia. *J Bone Miner Res* 2011;26(7):1381-1388.
4. **Connelly M, Neville K.** Comparative prospective evaluation of the responsiveness of single-item pediatric pain-intensity self-report scales and their uniqueness from negative affect in a hospital setting. *J Pain* 2010;11(12):1451-1460.
5. **DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, et al.** Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. *J Clin Epidemiol* 2011;64(7):794-804.
6. **European Medicines Agency (EMA).** Crysvida: European public assessment report EMEA/H/C/004275/0000 [online]. 14.12.2017. Amsterdam (NED): EMA; 2017. [Zugriff: 17.10.2019]. URL: https://www.ema.europa.eu/documents/assessment-report/crysvida-epar-public-assessment-report_en.pdf.
7. **Geiger R, Strasak A, Trembl B, Gasser K, Kleinsasser A, Fischer V, et al.** Six-minute walk test in children and adolescents. *J Pediatr* 2007;150(4):395-399, 399 e391-392.
8. **Gemeinsamer Bundesausschuss (G-BA).** Amendment 2/2 zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO; Wirkstoff: Burosumab [online]. Berlin (GER): G-BA; 2018. [Zugriff: 30.10.2019]. URL: https://www.g-ba.de/downloads/92-975-2502/2018-10-04_Amendment-2_Burosumab_D-349.pdf.
9. **Gemeinsamer Bundesausschuss (G-BA).** Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO; Wirkstoff: Burosumab [online]. Berlin (GER): G-BA; 2018. [Zugriff: 30.10.2019]. URL: https://www.g-ba.de/downloads/92-975-2392/2018-04-15_Nutzenbewertung-G-BA_Burosumab-D-349.pdf.
10. **Gemeinsamer Bundesausschuss (G-BA).** Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Burosumab vom 4. Oktober 2018 [online]. Berlin (GER): G-BA; 2018. [Zugriff: 30.10.2019]. URL: https://www.g-ba.de/downloads/40-268-5295/2018-10-04_AM-RL-XII_Burosumab_D-349_TrG.pdf.

11. **HealthMeasures.** List of pediatric measures: available PROMIS measures for pediatric self-report (ages 8-17) and parent proxy report (ages 5-17) [online]. 15.12.2017. Chicago (USA): Northwestern University; 2017. [Zugriff: 07.09.2018]. URL: <http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-pediatric-measures>.
12. **Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B.** The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. *Pain* 2001;93(2):173-183.
13. **Imel EA, Glorieux FH, Whyte MP, Munns CF, Ward LM, Nilsson O, et al.** Burosumab versus conventional therapy in children with X-linked hypophosphataemia: a randomised, active-controlled, open-label, phase 3 trial. *Lancet* 2019;393(10189):2416-2427.
14. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).** Allgemeine Methoden; Vers. 5.0 [online]. 10.07.2017. Köln (GER): IQWiG; 2017. [Zugriff: 27.11.2017]. URL: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.
15. **International Association for the Study of Pain (IASP).** Faces Pain Scale - Revised (FPS-R) [online]. Washington (USA): IASP; 2004. [Zugriff: 23.08.2018]. URL: https://s3.amazonaws.com/rdcms-iasp/files/production/public/Content/ContentFolders/Resources2/FPSR/facepainscale_english_eng-au-ca.pdf.
16. **Irwin DE, Gross HE, Stucky BD, Thissen D, DeWitt EM, Lai JS, et al.** Development of six PROMIS pediatrics proxy-report item banks. *Health Qual Life Outcomes* 2012;10:22.
17. **Irwin DE, Stucky BD, Thissen D, Dewitt EM, Lai JS, Yeatts K, et al.** Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. *Qual Life Res* 2010;19(4):585-594.
18. **Irwin DE, Varni JW, Yeatts K, DeWalt DA.** Cognitive interviewing methodology in the development of a pediatric item bank: a patient reported outcomes measurement information system (PROMIS) study. *Health Qual Life Outcomes* 2009;7:3.
19. **Kashikar-Zuck S, Carle A, Barnett K, Goldschneider KR, Sherry DD, Mara CA, et al.** Longitudinal evaluation of patient-reported outcomes measurement information systems measures in pediatric chronic pain. *Pain* 2016;157(2):339-347.
20. **Kyowa Kirin.** Crysvida 10/20/30 mg Injektionslösung [online]. Berlin: 01.2019. [Zugriff: 19.11.2019]. URL: <http://www.fachinfo.de>.
21. **Kyowa Kirin.** Dossier zur Nutzenbewertung gemäß § 35a SGB V: Burosumab. Modul 4 A X-chromosomale Hypophosphatämie (XLH): Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 2019.
22. **Laboratories ATSCoPSfCPF.** ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002;166(1):111-117.
23. **Lai JS, Stucky BD, Thissen D, Varni JW, DeWitt EM, Irwin DE, et al.** Development and psychometric properties of the PROMIS((R)) pediatric fatigue item banks. *Qual Life Res* 2013;22(9):2417-2427.

24. **Linglart A, Bioso-Duplan M, Briot K, Chaussain C, Esterle L, Guillaume-Czitrom S, et al.** Therapeutic management of hypophosphatemic rickets from infancy to adulthood. *Endocr Connect* 2014;3(1):R13-30.
25. **Saris-Baglana RN, DeRosa MA, Raczek AE, Bjorner JB, Ware JE.** Development, validation, and norming of the SF-10 for Children Health Survey. *Qual Life Res* 2006;15(Suppl 1):A145.
26. **Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG.** Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. *Cardiopulm Phys Ther J* 2013;24(3):21-29.
27. **Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG.** Triangulating Clinically Meaningful Change in the Six-minute Walk Test in Individuals with Chronic Heart Failure: A Systematic Review. *Cardiopulm Phys Ther J* 2012;23(3):5-15.
28. **Thissen D, Liu Y, Magnus B, Quinn H, Gipson DS, Dampier C, et al.** Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method. *Qual Life Res* 2016;25(1):13-23.
29. **Ultragenyx Pharmaceutical.** Clinical evaluator reference manual. Protocol: UX023-CL301; a randomized, open-label, phase 3 study to assess the efficacy and safety of KRN23 versus oral phosphate and active vitamin D treatment in pediatric patients with X-linked Hypophosphatemia (XLH); manual version 1.2 [unveröffentlicht]. 15. July 2016.
30. **Ultragenyx Pharmaceutical.** Patient-reported outcome measures: clinical outcome assessment dossier; The Pediatric Orthopaedic Society of North America Pediatric Outcomes Data Collection Instrument (POSNA-PODCI), National Institutes of Health Patient Reported Outcomes Measurement Information System (PROMIS), Faces Pain Scale - Revised (FPS-R) and 6-Minute Walk Test (6MWT) for pediatric patients with X-linked hypophosphatemia; Version 5.0 [unveröffentlicht]. 2017.
31. **Ultragenyx Pharmaceutical.** A randomized, open-label, phase 3 study to assess the efficacy and safety of KRN23 versus oral phosphate and active vitamin D treatment in pediatric patients with X-linked hypophosphatemia (XLH); KRN23 (UX023-CL301); clinical study protocol amendment 1 [unveröffentlicht]. 2017.
32. **Ultragenyx Pharmaceutical.** A randomized, open-label, phase 3 study to assess the efficacy and safety of KRN23 versus oral phosphate and active vitamin D treatment in pediatric patients with X-linked hypophosphatemia (XLH); KRN23 (UX023-CL301); Clinical study report - Week 64 [unveröffentlicht]. 2019.
33. **Ultragenyx Pharmaceutical.** A randomized, open-label, phase 3 study to assess the efficacy and safety of KRN23 versus oral phosphate and active vitamin D treatment in pediatric patients with X-linked hypophosphatemia (XLH); KRN23 (UX023-CL301); Statistical analysis plan; Version 1.0 [unveröffentlicht]. 2018.
34. **Varni JW, Magnus B, Stucky BD, Liu Y, Quinn H, Thissen D, et al.** Psychometric properties of the PROMIS (R) pediatric scales: precision, stability, and comparison of different scoring and administration options. *Qual Life Res* 2014;23(4):1233-1243.

35. **Varni JW, Stucky BD, Thissen D, Dewitt EM, Irwin DE, Lai JS, et al.** PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. *J Pain* 2010;11(11):1109-1119.
36. **Varni JW, Thissen D, Stucky BD, Liu Y, Gorder H, Irwin DE, et al.** PROMIS(R) Parent Proxy Report Scales: an item response theory analysis of the parent proxy report item banks. *Qual Life Res* 2012;21(7):1223-1240.
37. **Varni JW, Thissen D, Stucky BD, Liu Y, Magnus B, Quinn H, et al.** PROMIS(R) Parent Proxy Report Scales for children ages 5-7 years: an item response theory analysis of differential item functioning across age groups. *Qual Life Res* 2014;23(1):349-361.