

# 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**

Dossierbewertung vom 16. Juli 2018

Datum des Amendments: 13. September 2018

## Inhaltsverzeichnis

Tabellenverzeichnis.....	3
Abkürzungsverzeichnis.....	4
1. Hintergrund.....	5
2. Fragestellung.....	5
3. Studienbewertung.....	6
3.1. Liste der verwendeten Quellen.....	6
3.2. Studiencharakteristika.....	6
3.3. Endpunkte.....	12
3.3.1. In der Bewertung berücksichtigte Endpunkte.....	12
3.3.2. Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz.....	14
3.4. Verzerrungspotential auf Studien- und Endpunktebene.....	26
3.5. Ergebnisse.....	26
3.5.1. Charakterisierung der Studienpopulation.....	26
3.5.2. Mortalität.....	29
3.5.3. Morbidität.....	30
3.5.4. Lebensqualität.....	32
3.5.5. Sicherheit.....	33
4. Fazit.....	37
Referenzen.....	39

## Tabellenverzeichnis

Tabelle 1:	Charakterisierung der Studie UX023-CL301 .....	6
Tabelle 2:	Im Amendment berücksichtigte Endpunkte der Studie UX023-CL301 .....	12
Tabelle 3:	Beschreibung der Endpunkte der Kategorie Morbidität .....	14
Tabelle 4:	Beschreibung der Endpunkte der Kategorie Sicherheit .....	24
Tabelle 5:	Verzerrungspotential auf Studienebene .....	26
Tabelle 6:	Allgemeine Angaben zum Patientenfluss der Studie UX023-CL301 .....	26
Tabelle 7:	Charakterisierung der Studienpopulation von UX023-CL301 zu Baseline .....	27
Tabelle 8:	Exposition mit der Studienmedikation bis zum Datenschnitt am 12. Februar 2018 .....	28
Tabelle 9:	Begleitmedikationen zum Datenschnitt am 12. Februar 2018 bei > 30 % der Patienten in einer der Behandlungsgruppen der Studie UX023-CL301 .....	29
Tabelle 10:	Primärer Endpunkt: RGI-C-Gesamtwert zu Woche 40 in Studie UX023-CL301 .....	30
Tabelle 11:	Veränderung in der Stehhöhe bzw. Liegelänge zu Woche 40 in Studie UX023-CL301 ...	30
Tabelle 12:	Veränderung der Laufstrecke im 6MWT zu Woche 40 in Studie UX023-CL301, Patienten $\geq$ 5 Jahre der ITT-Population .....	31
Tabelle 13:	PROMIS-Scores zu Woche 40 in Studie UX023-CL301, Patienten $\geq$ 5 Jahre der ITT-Population .....	31
Tabelle 14:	Veränderung der akuten Schmerzintensität mittels FPS-R zu Woche 40 in Studie UX023-CL301, Patienten $\geq$ 5 Jahre der ITT-Population .....	32
Tabelle 15:	Zusammenfassung der UE zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301 .....	33
Tabelle 16:	UE mit Inzidenz $\geq$ 10 % in mindestens einem Studienarm zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301 .....	33
Tabelle 17:	UE mit CTCAE-Grad $\geq$ 3 zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301 .....	35
Tabelle 18:	SUE zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301 .....	35
Tabelle 19:	UE von besonderem Interesse zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301 .....	36
Tabelle 20:	Zusammenfassende Darstellung der Ergebnisse der Studie UX023-CL301 .....	37

## Abkürzungsverzeichnis

6MWT	6-Minuten-Gehtest (Six-Minute Walk Test)
ALP	Alkalische Phosphatase
ANCOVA	Kovarianzanalyse (Analysis of Covariance)
BALP	Knochenspezifische Alkalische Phosphatase
BMI	Body-Mass-Index
BOT-2	Bruininks-Oseretsky Test of Motor Proficiency – 2 <sup>nd</sup> Edition
COA	Clinical Outcome Assessment
CTCAE	Common Terminology Criteria for Adverse Events
FPS-R	Faces Pain Scale – Revised
G-BA	Gemeinsamer Bundesausschuss
GEE	Generalized Estimating Equation
GFR	glomeruläre Filtrationsrate
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
KI	Konfidenzintervall
MedDRA	Medical Dictionary for Regulatory Activities
MW	Mittelwert
PHEX	Phosphate regulating endopeptidase homolog X-linked
PRO	Patient-Reported Outcome/s
PROMIS	Patient-Reported Outcomes Measurement Information System
pU	pharmazeutischer Unternehmer
RCT	randomisierte kontrollierte Studie/n (Randomized Controlled Trial/s)
RGI-C	Radiographic Global Impression of Change
RSS	Rickets Severity Scale
SAP	statistischer Analyseplan
SD	Standardabweichung (Standard Derivation)
SE	Standardfehler (Standard Error)
SF-10	Short Form Health Survey-10 for Children
SGB	Sozialgesetzbuch
SUE	schwerwiegende/s unerwünschte/s Ereignis/se
TmP	maximale tubuläre Rückresorption von Phosphat
UE	unerwünschte/s Ereignis/se
VerfO	Verfahrensordnung des G-BA
XLH	X-chromosomale Hypophosphatämie (X-linked hypophosphatemia)

## 1. 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.

In die Bewertung werden gemäß Zulassungsstatus 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 [8].

Der pharmazeutische Unternehmer (pU) hatte am 15. April 2018 ein Dossier zur Nutzenbewertung gemäß § 35a SGB V beim Gemeinsamen Bundesausschuss (G-BA) eingereicht. Als bewertungsrelevante Studie wurden die Phase-II-Studien UX023-CL201 und UX023-CL205 für die Nutzenbewertung herangezogen. Dabei handelt es sich um multizentrische Studien ohne Kontrollarm, in der die Wirksamkeit und Sicherheit von Burosumab bei Kindern zwischen 1–4 Jahren bzw. 5–12 Jahren untersucht wird.

Die Nutzenbewertung wurde am 16. Juli 2018 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Eine mündliche Anhörung fand am 21. August 2018 statt, in der offene Fragen zu den in den schriftlichen Stellungnahmen dargelegten Daten und Argumentationen erörtert wurden.

### Phase-III-Studie UX023-CL301

Der pU hat im Stellungnahmeverfahren den ersten Datenschnitt der Studie UX023-CL301 eingereicht, der zum Zeitpunkt der Dossiereinreichung am 15. April 2018 noch nicht verfügbar war. Studienprotokoll und statistischer Analyseplan (SAP) der Studie UX023-CL301 wurden bereits mit Dossiereinreichung vorgelegt. Der Unterausschuss Arzneimittel hat im Nachgang zu der mündlichen Anhörung eine ergänzende Bewertung der eingereichten Unterlagen der Studie UX023-CL301 beauftragt.

## 2. Fragestellung

Mit diesem Amendment werden die am 6. August 2018 durch den pU im schriftlichen Stellungnahmeverfahren eingereichten Daten der Studie UX023-CL301 sowie die im Anschluss an die mündliche Anhörung am 21. August 2018 eingereichten Unterlagen zum Wirkstoff Burosumab bewertet.

### 3. Studienbewertung

#### 3.1. Liste der verwendeten Quellen

Zur Bestimmung des Ausmaßes des Zusatznutzens für Burosumab wurden folgende Quellen für dieses Amendment herangezogen:

- Schriftliche Stellungnahme des pU vom 06.08.2018 [18],
- Studienbericht, SAP und Studienprotokoll der Studie UX023-CL301 [24,25,26],
- Clinical Outcome Assessment (COA-) Dossier [23].

#### 3.2. Studiencharakteristika

Tabelle 1: Charakterisierung der Studie UX023-CL301

Charakteristikum	Beschreibung
<b>Design</b>	<p>Die Studie UX023-CL301 ist eine multizentrische, randomisierte, 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 <math>\leq</math> 12 Jahren. Primäres Zielkriterium der Studie ist die Verbesserung der Rachitis basierend auf dem RGI-C-Global-Score von Baseline zu Woche 40.</p> <p>Die Randomisierung erfolgte stratifiziert nach folgenden Faktoren:</p> <ul style="list-style-type: none"> <li>• Rachitisschwere (RSS <math>\leq</math> 2,5 vs. <math>&gt;</math> 2,5)</li> <li>• Alter (<math>&lt;</math> 5 vs. <math>\geq</math> 5 Jahre)</li> <li>• Region (Japan vs. Rest der Welt)</li> </ul> <p>Die Studie besteht aus einer Screeningperiode (einschließlich einer einwöchigen Phosphat- und Vitamin-D-Wash-Out-Phase), einer 64-wöchigen kontrollierten Behandlungsperiode und einer bis zu 76-wöchigen Extentionsperiode (nur in Europa, Nordamerika und Australien), in der alle Patienten Burosumab erhalten (s. Abbildung 1). Die Studie ist noch nicht abgeschlossen. Die Ergebnisse der primären Analyse (Woche 40, s. u. Zeile „Statistische Methoden“) liegen vor.</p>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Mädchen und Jungen im Alter von 1 bis 12 Jahren mit radiologischem Nachweis einer Rachitis, definiert als RSS-Gesamtscore von <math>\geq</math> 2,0</li> <li>• PHEX-Mutation (oder Variante mit unklarer Bedeutung) beim Patienten oder einem direkt verwandten Familienmitglied mit entsprechender X-chromosomaler Vererbung</li> <li>• Biochemische Befunde, die mit XLH assoziiert sind: Serum-Phosphat <math>&lt;</math> 3,0 mg/dl (0,97 mmol/l)</li> <li>• Serum-Kreatinin unterhalb der altersadjustierten oberen Normalbereichsgrenze</li> <li>• Serum-25(OH)D über dem unteren Normwert (<math>\geq</math> 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 der Proband sonst alle anderen Einschlusskriterien erfüllt, erfolgt ein Rescreening nach einem Minimum von 7 Tagen)</li> <li>• Orale Einnahme von Phosphat und aktivem Vitamin D für <math>\geq</math> 12 aufeinanderfolgende Monate (bei Kindern <math>\geq</math> 3 Jahre) oder für <math>\geq</math> 6 aufeinanderfolgende Monate (bei Kindern <math>&lt;</math> 3 Jahre) bis 7 Tage vor der Randomisierung</li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Tanner-Stadium von 4 oder höher in einem der folgenden Merkmale: Genitalien, Brust oder Schamhaare</li> <li>• Perzentil der Körperhöhe <math>&gt;</math> 50 basierend auf länderspezifischen Normen</li> <li>• Verwendung von Aluminiumhydroxid-Antazida, systemischen Kortikosteroiden, Acetazolamid und Thiaziden innerhalb von 7 Tagen vor Screening-Visite</li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>• Gegenwärtige oder frühere Verwendung von Leuprorelin, Triptorelin, Goserelin oder anderen Arzneimitteln, von denen bekannt ist, dass sie die Pubertät verzögern</li> <li>• Wachstumshormontherapie innerhalb von 12 Monaten vor Screening-Visite</li> <li>• Vorhandensein einer Nephrokalzinose im Nierenultraschall Grad 4 basierend auf folgender Skala: <ul style="list-style-type: none"> <li>○ 0: normal</li> <li>○ 1: schwacher hyperechogener Rand um die Markpyramiden</li> <li>○ 2: intensiverer echogener Rand mit Echos, mit schwacher Füllung der Pyramide</li> <li>○ 3: gleichmäßig intensive Echos in der ganzen Pyramide</li> <li>○ 4: Steinbildung: solitärer Echo-Fokus an der Spitze der Pyramide</li> </ul> </li> <li>• Geplante oder empfohlene orthopädische Operation einschließlich Osteotomie, Implantation oder Entfernung von Klammern (Staples), 8-Platten oder anderen orthopädischen Materialien innerhalb der ersten 40 Wochen der Studie</li> <li>• Hypo- oder Hyperkalzämie, definiert als Serum-Kalziumspiegel außerhalb der altersajustierten Normalwerte</li> <li>• Nachweis eines Hyperparathyreoidismus (PTH-Level 2,5 x ULN)</li> <li>• Einnahme von PTH-Suppressiva innerhalb von 2 Monaten vor Screening-Visite</li> <li>• Vorhandensein oder Vorgeschichte eines Zustandes des XLH-Patienten, den der Prüfarzt als ein hohes Risiko für eine schlechte Compliance oder einen möglichen Studienabbruch einstuft</li> <li>• Vorhandensein einer Begleiterkrankung oder eines Gesundheitszustandes, welche/r die Teilnahme an der Studie beeinträchtigen würde</li> <li>• Vorgeschichte einer rezidivierenden Infektion oder Veranlagung zur Infektion oder einer bekannten Immunschwäche</li> <li>• Verwendung eines monoklonalen Antikörpers innerhalb von 90 Tagen vor der Screening-Visite oder vorherige allergische oder anaphylaktische Reaktionen auf einen monoklonalen Antikörper</li> <li>• Gegenwärtige oder frühere Überempfindlichkeit gegen Burosumab-Hilfsstoffe, die nach Einschätzung des Prüfarztes mit einem erhöhten Risiko für UE einhergehen</li> </ul>
<b>Intervention und Zahl der Patienten</b>	<p><b>Interventionsgruppe:</b> Burosumab (N=29)</p> <ul style="list-style-type: none"> <li>• Dosierung: <ul style="list-style-type: none"> <li>○ 0,8 mg/kg Körpergewicht (KG) alle 2 Wochen (Q2W)</li> <li>○ Dosiserhöhung auf 1,2 mg/kg KG basierend auf Nüchtern-Serum-Phosphat-Wert möglich<sup>1)</sup></li> </ul> </li> <li>• Applikation: Subkutane Injektion in Bauch, Oberarm oder Oberschenkel (Injektionsstelle wechselt bei jeder Injektion). Übersteigt die Dosis ein Volumen von 1,5 ml, sollte die Dosis an zwei Injektionsstellen verabreicht werden.</li> <li>• Keine orale Supplementation mit Phosphat und aktivem Vitamin D während der Studie.</li> </ul> <p><b>Kontrollgruppe:</b> Orales Phosphat und aktives Vitamin D (N=32)</p> <ul style="list-style-type: none"> <li>• Dosierung: individuell nach Ermessen des Prüfarztes auf der Grundlage US-amerikanischer und europäischer Leitlinien <ul style="list-style-type: none"> <li>○ US-Leitlinie: Calcitriol 20–30 ng/kg/Tag aufgeteilt in 2 bis 3 Dosen und Phosphat in einer Elementarphosphordosis von 20–40 mg/kg/Tag aufgeteilt in 3 bis 5 Dosen [5];</li> <li>○ Europäische Leitlinie: Alfacalcidol 1–2 µg/Tag und Phosphat 45–70 mg/kg/Tag aufgeteilt in 3 bis 4 Dosen [20].</li> </ul> </li> </ul> <p>Die Behandlungsdauer beträgt in der Interventions- und Kontrollgruppe 64 Wochen. Im Anschluss an die kontrollierte Behandlungsperiode können die Patienten der Interventions- und Kontrollgruppe (nach 7-tägiger Phosphat- und Vitamin-D-Wash-Out-Phase) in einer unkontrollierten Extensionsperiode bis zu 76 weitere Wochen mit Burosumab behandelt werden.</p>

Charakteristikum	Beschreibung
	<p><b>Nicht erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Aluminiumhydroxid-Antazide, systemische Kortikosteroide, Acetazolamid oder Thiazid</li> <li>• Wachstumshormone</li> <li>• Parathormon-Suppressiva</li> <li>• Jede monoklonale Antikörper-Therapie (außer Studienmedikation)</li> <li>• Leuprorelin, Triptorelin, Goserelin oder andere Medikamente, von denen bekannt ist, dass sie die Pubertät verzögern</li> </ul> <p>Patienten der Burosumab-Gruppe durften während der gesamte Studie (kontrollierte Behandlungs- und Extensionsperiode) kein orales Phosphat und kein aktives Vitamin D einnehmen. Patienten der Kontrollgruppe durften während der Extensionsperiode kein orales Phosphat und kein aktives Vitamin D einnehmen.</p> <p><b>Erlaubte Begleitmedikation</b></p> <p>Notwendige Begleitmedikation (abgesehen von der nicht erlaubten Begleitmedikation) war gestattet, wobei keine signifikanten Änderungen der Medikation während der Studie vorgenommen werden sollten.</p>
<b>Ort und Zeitraum der Durchführung</b>	<p><b>Ort der Durchführung</b></p> <p>16 Studienzentren: USA (5), Kanada (3), Japan (3), Großbritannien (2), Schweden (1), Australien (1), Korea (1).</p> <p><b>Studiendauer</b></p> <ul style="list-style-type: none"> <li>• Einschluss erster Patient: 8. September 2016</li> <li>• Einschluss letzter Patient: keine Angabe</li> <li>• Aktueller Datenschnitt: 12. Februar 2018 (Datenschnitt der primären Analyse) <ul style="list-style-type: none"> <li>○ Primäre Analyse fand statt, nachdem der letzte Patient Studienwoche 40 durchlaufen hatte</li> <li>○ Zu diesem Zeitpunkt befanden sich alle Patienten noch in der Studie</li> </ul> </li> <li>• Geplantes Studienende: September 2018 / Juli 2019 (Ende der Extensionsperiode)</li> <li>• Weitere geplante Datenschnitte: <ul style="list-style-type: none"> <li>○ Studienwoche 64 (Ende der kontrollierten Behandlungsperiode)</li> <li>○ Studienwoche 140 (Ende der Extensionsperiode)</li> </ul> </li> </ul>
<b>Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte gemäß Studienprotokoll</b>	<p><b>Primärer Endpunkt</b></p> <p>Veränderung im RGI-C-Global-Score von Baseline zu Woche 40</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Anteil der Patienten mit RGI-C-Gesamtwert <math>\geq + 2,0</math> an Woche 40 und Woche 64</li> <li>• Veränderung der Rachitis basierend auf dem RGI-C-Gesamtwert zu Woche 64</li> <li>• Veränderung im RSS-Gesamtwert von Baseline zu Woche 40 und Woche 64</li> <li>• Veränderung der Skelettanomalien der unteren Extremitäten basierend auf dem RGI-C-Bein-Score zu Woche 40 und 64</li> <li>• Veränderung in der Stehhöhe (oder Körperlänge im Liegen bei Kindern &lt; 2 Jahre) von Baseline zu Woche 24, 40 und 64</li> <li>• Veränderung im altersnormierten Z-Score der Körperhöhe zu Woche 24, 40 und 64</li> <li>• Veränderung in der Wachstumsgeschwindigkeit</li> <li>• Schmerz, Fatigue und körperliche Funktion: Veränderung in den PROMIS-Domänenwerten „Pediatric Pain Interference“, „Physical Function Mobility“ und „Fatigue“ (bei Patienten <math>\geq 5</math> Jahre) zu Woche 24, 40 und 64</li> <li>• Schmerzintensität: Veränderung im FPS-R (bei Patienten <math>\geq 5</math> Jahre) zu Woche 24, 40 und 64</li> <li>• Gehfähigkeit: Veränderung der 6MWT-Gehstrecke und Anteil der erwarteten 6MWT-Gehstrecke (bei Patienten <math>\geq 5</math> Jahre) zu Woche 24, 40 und 64</li> </ul>

Charakteristikum	Beschreibung
	<p><b>Explorative Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Gesundheitsbezogene Lebensqualität mittels SF-10 (für Kinder ab 5 Jahren) zu Woche 24, 40 und 64</li> <li>• Zahnärztliche Beurteilung: Anzahl der Ereignisse bzgl. Karies, verzögerter Zahndurchbruch, Schmelzhypoplasie, Zahnabszess und Gingivitis</li> </ul> <p><b>Pharmakodynamische Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Phosphat, 1,25(OH)<sub>2</sub>D im Serum</li> <li>• Phosphat im Urin, Renale Phosphatschwelle (TmP/GFR) und prozentuale tubuläre Phosphatrückresorption (TRP)</li> <li>• Knochenmarker</li> </ul> <p><b>Pharmakokinetik</b> Burosumab im Serum</p> <p><b>Sicherheit</b> Inzidenz der UE und SUE</p>
<b>Statistische Methoden</b>	<p><b>Fallzahlplanung</b></p> <ul style="list-style-type: none"> <li>• Basierend auf mittlerem RGI-C-Gesamtwert (primärer Endpunkt)</li> <li>• Annahmen: <ul style="list-style-type: none"> <li>○ mittlerer RGI-C-Gesamtwert von 1,80 in Burosumab-Gruppe und 1,40 in Kontrollgruppe</li> <li>○ Standardabweichung: 0,50</li> <li>○ Drop-out-Rate: 10 %</li> </ul> </li> <li>• Power: 80 %</li> <li>• zweiseitiger <math>\alpha</math>-Wert: 0,05</li> <li>• benötigte Gesamtstichprobengröße: 60 Patienten (30 pro Behandlungsgruppe)</li> </ul> <p><b>Analysepopulationen</b></p> <p><u>Full Analysis Set (FAS)<sup>2)</sup>:</u> Alle randomisierten Patienten mit mindestens einer Dosis der zugeteilten Studienmedikation und mindestens einer Post-Baseline-Bewertung. Die Patienten werden entsprechend ihrer randomisierten Behandlung ausgewertet.</p> <p>Full Analysis-Subset der Patienten <math>\geq 5</math> Jahre: Alle Patienten des FAS mit einem Alter <math>\geq 5</math> Jahre zu Baseline (Stratifizierungsfaktor bei Randomisierung). Die Auswertungen der Endpunkte PROMIS, 6MWT und FPS-R erfolgen mit diesem Subset.</p> <p><u>Safety Analysis Set</u> Alle randomisierten Patienten mit mindestens einer Dosis der zugeteilten Studienmedikation. Die Patienten werden entsprechend der tatsächlich erhaltenen Behandlung ausgewertet.</p> <p><b>Statistische Analysen</b></p> <p>Statistische Tests wurden zweiseitig auf dem Signifikanzniveau <math>\alpha = 0,05</math> und mit 95%-KI durchgeführt. Eine Adjustierung für multiples Test wurde nicht durchgeführt.</p> <p>Der primäre Endpunkt RGI-C, als Änderung zwischen Baseline und Woche 40 und als Unterschied dieser Veränderung zwischen den Studienarmen, wurde mittels ANCOVA untersucht. Behandlungsgruppe, Alter, Stratifizierungsfaktoren und RSS zu Baseline wurden als Adjustierungsfaktoren berücksichtigt.</p> <p>Anthropometrische Maßzahlen wurden als Z-Scores und Perzentile dargestellt.</p> <p>Für Endpunkte mit wiederholten Messungen (6MWT, PROMIS, FPS-R, Anthropometrie) während der Studienphase wurden GEE-Modelle herangezogen. Es wurde für Behandlungsgruppe, Studiervisite, Interaktionsterm aus Behandlungsgruppe und Studiervisite, Stratifizierungsfaktoren, RSS zu Baseline und Alter adjustiert.</p>

Charakteristikum	Beschreibung
	Sensitivitätsanalysen zur Untersuchung des Einflusses von fehlenden Werten wurden für den primären Endpunkt RGI-C mit den Ersetzungsverfahren last observation carried forward und multiple Imputation geplant.

<sup>1)</sup> Anpassung der Burosumab-Dosis:

a) Kriterien zur Dosiserhöhung:

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

b) Wenn das Serum-Phosphat über die obere Grenze des Normalbereichs steigt, werden die nachfolgenden Dosen zunächst ausgesetzt. Nach Ausschluss anderer Ursachen des erhöhten Serum-Phosphats wird die Burosumab-Behandlung mit der Hälfte der zuletzt erhaltenen Dosis unter Überwachung der Serum-Phosphat-Werte fortgesetzt.

<sup>2)</sup> Alle randomisierten Patienten der Studie UX023-CL301 erhielten mindestens eine Dosis der Studienmedikation, sodass die FAS-Population dem ITT-Prinzip entspricht.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; ALP: Alkalische Phosphatase; ANCOVA: Kovarianzanalyse; BOT-2: Bruininks-Oseretsky Test of Motor Proficiency – 2<sup>nd</sup> Edition; FAS: Full Analysis Set; FPS-R: Faces Pain Scale – Revised; ITT: Intention-to-Treat; KI: Konfidenzintervall; PHEX: Phosphate regulating endopeptidase homolog X-linked; PROMIS: Patient-Reported Outcomes Measurement Information System; PTH: Parathormon; Q2W: alle 2 Wochen; RGI-C: Radiographic Global Impression of Change; RSS: Rickets Severity Scale; (S)UE: (schwerwiegende/s) unerwünschte/s Ereignis/se; TmP: maximale tubuläre Rückresorption von Phosphat; XLH: X-chromosomale Hypophosphatämie.

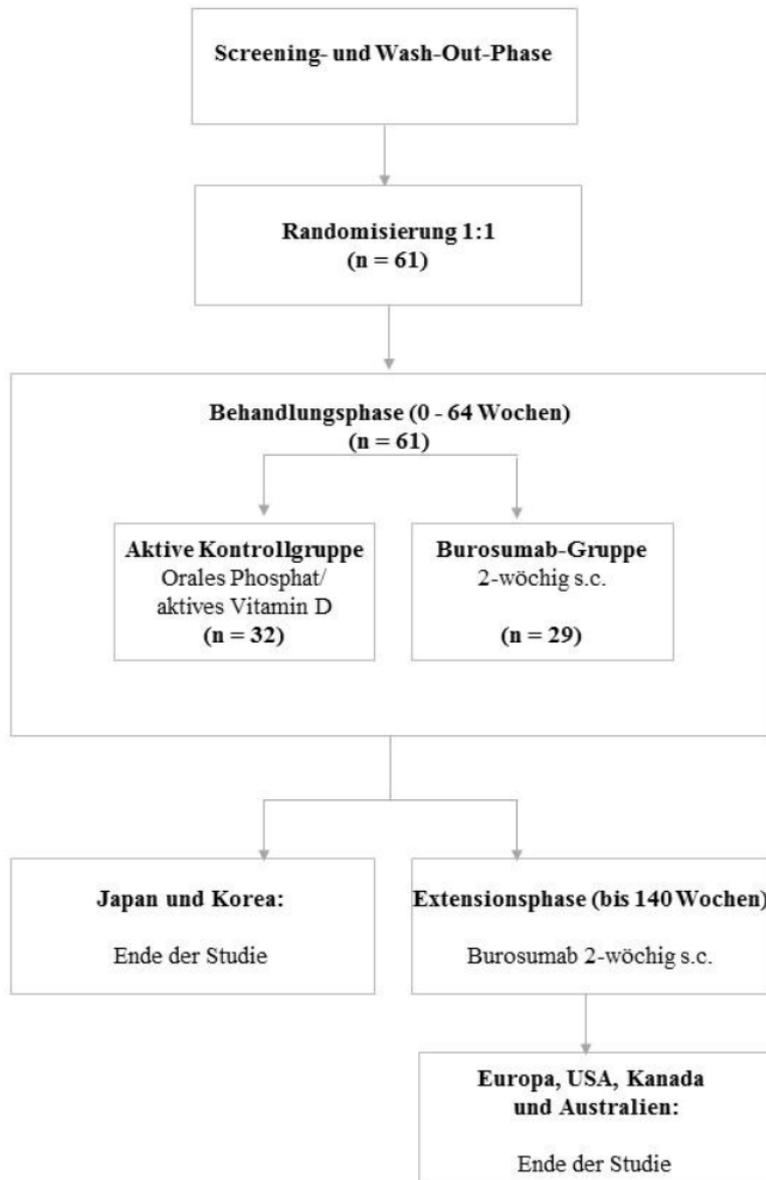


Abbildung 1: Patientenfluss in der Studie UX023-XL301 (Behandlungsphase noch nicht abgeschlossen)

### Protokolländerungen

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

- Einführung der einarmigen Extensionsphase (bis zu 76 Wochen) für Patienten in Europa, den USA, Kanada und Australien;
- Änderung der Nachbeobachtung der Sicherheitsendpunkte: Einschränkung der 12-wöchigen Nachbeobachtung auf lediglich diejenigen Patienten, die nicht anderweitig mit Burosumab therapiert werden (z. B. durch Teilnahme an einer anderen Burosumab-Studie, s. Tabelle 4);
- Erhöhung des Anteils von weiblichen Studienteilnehmern auf 70 %;

- Präzisierung von Ein- und Ausschlusskriterien;
- Änderung der PROMIS-Version von 1.0 auf 2.0;
- Definition des Endpunkts „Zahnärztliche Beurteilung“ als Karies, verzögerter Zahndurchbruch, Schmelzhypoplasie, Zahnabszess und Gingivitis.

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

Die Änderung der Version des PROMIS (Patient-Reported Outcomes Measurement Information System) 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 Darstellung im Nachgang der mündlichen Anhörung wurde in der Studie nur Version 2.0 des PROMIS 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 in Abschnitt 0 beschrieben.

### 3.3. Endpunkte

#### 3.3.1. In der Bewertung berücksichtigte Endpunkte

Tabelle 2: Im Amendment berücksichtigte Endpunkte der Studie UX023-CL301

Endpunkt	Kategorie	Vom pU als patientenrelevant eingestuft	In der Bewertung berücksichtigt
Rachitissymptomatik mittels RSS	Morbidität	•	-
Rachitissymptomatik mittels RGI-C		•	o <sup>1)</sup>
Knochenmineralisierung (ALP)		•	-
Serum-Phosphat		•	-
1,25-Dihydroxy-Vitamin D		•	-
TmP, GFR		•	-
Anthropometrische Parameter (Körpergröße)		•	•
Gehfähigkeit (6MWT)		•	•
PROMIS-Domänen: <ul style="list-style-type: none"> <li>• Pediatric Pain Interference</li> <li>• Physical Function Mobility</li> <li>• Fatigue</li> </ul>		•	•
Schmerzintensität (FPS-R)		•	•
Zahnärztliche Beurteilung		x	-

Endpunkt	Kategorie	Vom pU als patientenrelevant eingestuft	In der Bewertung berücksichtigt
SF-10	Gesundheitsbezogene Lebensqualität	•	-2)
UE, SUE	Sicherheit	•	•

- Endpunkt wird als bewertungsrelevant eingestuft und in der Studienbewertung berücksichtigt.
- Endpunkt wird nicht als bewertungsrelevant eingestuft aber ergänzend in der Studienbewertung dargestellt.
- Endpunkt wird nicht als bewertungsrelevant eingestuft.
- x Entgegen der Einschätzung des pU wird der Endpunkt als bewertungsrelevant eingestuft und in der Studienbewertung berücksichtigt.

<sup>1)</sup> Rachitissymptomatik mittels RGI-C ist der primäre Endpunkt der Studie UX023-CL301 und wird für diese Studie ergänzend dargestellt.

<sup>2)</sup> Aufgrund von Einschränkungen in den psychometrischen Eigenschaften werden die Ergebnisse in der Nutzenbewertung nicht dargestellt (s. Nutzenbewertung D-349 [10]).

Abkürzungen: 6MWT: 6-Minuten-Gehtest; ALP: Alkalische Phosphatase; FPS-R: Faces Pain Scale – Revised; GFR: glomeruläre Filtrationsrate; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; RGI-C: Radiographic Global Impression of Change; RSS: Rickets Severity Scale; SF-10: Short Form Health Survey-10 for Children; (S)UE: (schwerwiegende/s) unerwünschte/s Ereignis/se; TmP: maximale tubuläre Rückresorption von Phosphat.

Zur Patientenrelevanz der Endpunkte Rachitissymptomatik mittels RGI-C (Radiographic Global Impression of Change) und RSS (Rickets Severity Scale), Knochenmineralisierung (alkalische Phosphatase [ALP]), Serum-Phosphat, 1,25-Dihydroxy-Vitamin D, TmP (maximale tubuläre Rückresorption von Phosphat), GFR (glomeruläre Filtrationsrate) und Wachstumsgeschwindigkeit wurden keine neuen Nachweise durch den pU vorgelegt, sodass eine Patientenrelevanz weiterhin unklar bleibt [10]. Als primärer Endpunkt der Studie werden die Ergebnisse der Erhebungen mittels RGI-C ergänzend dargestellt.

Die Einschätzung der psychometrischen Eigenschaften zum Instrument SF-10 (Short Form Health Survey-10 for Children) ist unverändert zur Nutzenbewertung [10], sodass die Ergebnisse nicht dargestellt werden.

### Zahnärztliche Beurteilung

Zahnärztliche Beurteilung ist ein kombinierter Endpunkt, definiert als Anzahl zahnmedizinischer Ereignisse bei Karies, verzögerter Durchbruch der Zähne, Zahnschmelzhypoplasie, Zahnabszesse und Gingivitis, die zu Studienbeginn und nach Verabreichung der Studienmedikation beurteilt wurden. Die Erhebung des Endpunktes erfolgte als zahnärztliche Untersuchung für die Sicherheitspopulation zu Woche 0 (Studienbeginn), den Wochen 4, 8, 16, 24, 32, 40, 52, und Woche 64 (Studienende) sowie zu Woche 74. Bei jedem Klinikbesuch wurden die Studienteilnehmer zudem aktiv befragt, ob seit der letzten Studiervisite die genannten zahnärztlichen Ereignisse aufgetreten sind. Separate zahnärztliche Beurteilungen sind nicht erfolgt und weitere zahnärztliche Ereignisse wurden als unerwünschtes Ereignis erfasst.

Die Erkrankung ist durch eine Beeinträchtigung des Knochenstoffwechsels und des Knochenapparates gekennzeichnet. Kinder mit XLH können schwere Zahndefekte aufweisen, die in Form von Spontaninfektionen infolge einer defekten Dentin- und Zahnschmelzhypoplasie auftreten, die zu Zahnabszessen führen. Zahnabszesse werden aufgrund der Symptomlast und möglicher Folgekomplikationen per se als patientenrelevant betrachtet. Zudem sind durch den Endpunkt neben Zahnabszessen weitere zahnmedizinische Ereignisse zusammengefasst, d. h. patientenrelevante mit nicht relevanten Ereignissen (z. B. verzögerter Zahndurchbruch). Die summarische

Darstellung der Ereignisse erscheint daher nicht sinnvoll. Aus der Operationalisierung des Endpunktes geht nicht hervor, wie die zahnärztliche Untersuchung erfolgt ist und ob beispielsweise regelhaft Röntgenbilder zur Beurteilung der nicht sichtbaren Zahnbestandteile angefertigt wurden. Individuelle Abweichungen abhängig vom Prüfarzt sind daher vorstellbar. Zudem wird nicht beschrieben, ob und wie Informationen aus zahnärztlichen Untersuchungen außerhalb der Studienvisiten übermittelt wurden. Eine Auswertung und Darstellung als unerwünschtes Ereignis nach Systemorganklassen wird zudem kritisch gesehen, da bei den Zahnproblemen nicht unterschieden werden kann, ob es sich um eine Nebenwirkung handelt oder diese mit der Symptomatik der Grunderkrankung einhergehen. Der Endpunkt wird aufgrund der dargelegten Unsicherheiten in der Operationalisierung und aufgrund der nicht patientenrelevanten Komponenten des zusammengesetzten Endpunkts nicht für die weitere Studienbewertung berücksichtigt.

### 3.3.2. Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz

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. Zur Beurteilung der Patientenrelevanz der vorgelegten Endpunkte wurden das Dossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ferner wurden eigene Recherchen zur Validität und Patientenrelevanz der Endpunkte durchgeführt. Das Ergebnis der Betrachtung der Endpunkte findet sich aufgeteilt nach Endpunktkategorien in Tabelle 3 und Tabelle 4.

*Tabelle 3: Beschreibung der Endpunkte der Kategorie Morbidität*

<b>Endpunktkategorie Morbidität</b>	
<b>Endpunkt</b>	<b>Veränderung der Rachitissymptomatik mittels RGI-C</b>
<b>Operationalisierung</b>	<p>Zur Einschätzung der Veränderung der Rachitissymptomatik wurde der Radiographic Global Impression of Change (RGI-C-) Score verwendet. Dabei handelt es sich um eine 7-stufige Ordinalskala (-3 bis +3), mit der Veränderungen der Rachitis über die Zeit anhand von Röntgenbildern bewertet werden.</p> <p>Drei pädiatrische Radiologen, die nicht an der Durchführung der Studie beteiligt waren (kein Zugang zum Studienprotokoll und zur Patientenidentifikation, keine Informationen zur Behandlung) führten verblindete Bewertungen von Handgelenk-, Knie- und Bein-Röntgenaufnahmen durch. Es stand ein Manual zur Bewertung zur Verfügung. Vor der Bewertung von Röntgenaufnahmen wurden die Radiologen in der Verwendung des RGI-C geschult, um einen Konsens über die Terminologie zur Beschreibung von XLH-bezogenen Röntgenanomalien zu erzielen. Die Bewertungen wurden unabhängig voneinander mithilfe eines elektronischen Datenerfassungssystems durchgeführt, wobei die Bewerter keine Möglichkeit hatten, Bilder zu diskutieren oder Bewertungen zu vergleichen; die Bewertungen konnten von den Bewertern nach der Einreichung nicht abgerufen oder geändert werden. Die Bewertung der Röntgenbildpaare (Baseline und Postbaseline) der jeweiligen Patienten erfolgte in zufälliger Reihenfolge.</p> <p>Zur Ermittlung des RGI-C wurde zunächst das Vorhandensein oder Nichtvorhandensein von XLH-bezogenen Anomalien (metaphysäre Aufhellung, metaphysäre/epiphysäre Trennung, metaphysäre Unschärfe, metaphysäre konkave Becherung) in der distalen Ulna und im Radius der beiden Handgelenke sowie in der proximalen Tibia, der proximalen Fibula und dem distalen Femur der beiden Knie für die Baseline-Aufnahmen bewertet. Durch anschließenden direkten Vergleich der Baseline-Aufnahmen mit den Post-Baseline-Aufnahmen erfolgte die Beurteilung einer Veränderung des radiologischen Erscheinungsbilds der oben genannten vier radiographischen Merkmale als unverändert, verringert, vergrößert oder neu aufgetreten.</p>

<b>Endpunktkategorie Morbidität</b>	
	<p>Basierend auf dieser Einschätzung aller sichtbaren Anomalien auf einem Röntgenbildpaar erfolgte die Bewertung der Veränderung auf einer 7-Punkte-Skala zur Ermittlung des RGI-C-Handgelenks- bzw. -Knie-Scores.</p> <p>7-Punkte-Skala:</p> <ul style="list-style-type: none"> <li>+3 sehr deutliche Verbesserung / (nahezu) vollständige Heilung</li> <li>+2 erhebliche Verbesserung/wesentliche Heilung</li> <li>+1 minimale Verbesserung</li> <li>0 keine Veränderung</li> <li>-1 minimale Verschlechterung</li> <li>-2 mittelstarke Verschlechterung</li> <li>-3 starke Verschlechterung</li> </ul> <p>Anschließend wurde ein RGI-C-Gesamtwert auf der Grundlage des Gesamteindrucks der beobachteten Veränderungen in den Handgelenk- und Knieaufnahmen unter Verwendung der 7-Punkte-Skala vergeben. Der RGI-C-Gesamtwert basiert auf einer globalen Einschätzung der Veränderung der Handgelenks- und Knieaufnahmen und wird nicht auf der Grundlage der bereits zugewiesenen RGI-C-Handgelenks- und Knie-Scores berechnet.</p> <p>Zusätzlich wurde basierend auf den Röntgenaufnahmen der Beine (stehend) ein RGI-C-Bein-Score ermittelt, der die Veränderung in Bezug auf Beindeformitäten abbilden soll. Der RGI-C-Bein-Score trägt nicht zum RGI-C-Gesamtwert bei.</p> <p>Für jeden RGI-C-Score (Handgelenk, Knie, Global, Bein) wurde der Durchschnitt der von den 3 unabhängigen Radiologen vergebenen Scores zur Analyse herangezogen.</p> <p><b>Erhebungszeitpunkte während der kontrollierten Behandlungsphase:</b> Radiographische Untersuchungen erfolg(t)en zu Screening/Baseline, Woche 40 und Woche 64 (bzw. bei vorzeitigem Abbruch, wenn innerhalb der letzten drei Monate keine Röntgenaufnahmen gemacht wurden).</p>
<b>Bewertung</b>	<p><b>Operationalisierung:</b> Die Operationalisierung ist mit Einschränkungen nachvollziehbar beschrieben. Unklarheiten bestehen hinsichtlich folgender Punkte:</p> <ul style="list-style-type: none"> <li>• Wie wurden die Stufen der Verschlechterung/Verbesserung (anhand der Manifestationszeichen) operationalisiert?</li> <li>• Welchen Maßnahmen wurden eingesetzt, um die Bestimmung des Gesamtwerts zu objektivieren?</li> </ul> <p><b>Validität:</b> Der pU führt in seiner schriftlichen Stellungnahme an, dass die „RGI-C-Skala [...] von pädiatrischen Radiologen mit Fachkenntnissen in der Skelettdysplasie entwickelt [wurde], die sich mit der Untersuchung von Handgelenks- und Knieaufnahmen von Kindern mit XLH befassten“. Literatur, die den Entwicklungsprozess beschreibt, konnte nicht identifiziert werden. Der pU hat eigene Untersuchungen der Intra- und Interraterreliabilität des RGI-C-Gesamtwerts basierend auf den Daten der XLH-Studien UX023-CL201 (Interventionsstudie; N=52) und einem zufällig ausgewählten Subset (N=30) der retrospektiven Beobachtungsstudie UX023-CL002 durchgeführt [3].</p> <p>Die Intraraterreliabilität des RGI-C-Gesamtwerts wurde durch eine erneute Bewertung der radiographischen Bildpaare (Baseline und Postbaseline) der Studien UX023-CL201 und UX023-CL002 durch die drei beteiligten Radiologen untersucht. Die Retest-Reliabilität des RGI-C-Gesamtwerts wird als ausreichend bewertet (weighted Kappa-Werte &gt; 0,5).</p> <p>Zur Bestimmung der Interraterreliabilität des RGI-C-Gesamtwerts wurden die Ergebnisse der ersten Bewertung sowie der Wiederholungsbewertung des Datensets der Intraraterreliabilitätsuntersuchung zwischen den drei Bewertern verglichen. Die Übereinstimmung wurde jeweils von 2 der 3 Radiologen verglichen, daraus ergaben sich 3 Bewerterpaare. Die Interraterreliabilität über alle drei Radiologen wurde nicht berichtet. Aufgrund der Abhängigkeit der ersten und zweiten Bewertung werden für die vorlie-</p>

<b>Endpunktkategorie Morbidität</b>	
	<p>gende Einschätzung der Reliabilität nur die Ergebnisse der ersten RGI-C-Bewertung herangezogen. Je nach Bewerterpaar variierte das gewichtete Kappa in der Studie UX023-CL201 von 0,3–0,4 und in UX023-CL002 von 0,4–0,6. Diese Daten deuten nur auf eine schwache bis moderate Übereinstimmung hin. Sie stärken die Vermutung, dass Schwächen in der Objektivität der globalen Einschätzung des RGI-C bestehen.</p> <p><b>Patientenrelevanz:</b> 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>Als rein radiologischer Endpunkt, der keine direkt durch den Patienten wahrnehmbare Symptomatik erfasst, ist der RGI-C nicht per se patientenrelevant. Daten, die die Validität des RGI-C als Surrogat für Morbidität im vorliegenden Anwendungsgebiet belegen, liegen nicht vor. Damit ist die Patientenrelevanz des RGI-C unklar.</p> <p>Da es sich um den primären Endpunkt der Studie UX023-CL301 handelt, werden die Ergebnisse trotz unklarer Patientenrelevanz und den beschriebenen Einschränkungen in der Objektivität ergänzend abgebildet.</p>
<b>Endpunkt</b>	<b>Anthropometrische Parameter (Körpergröße)</b>
<b>Operationalisierung</b>	<p>Das Wachstum wurde als Veränderung der Stehhöhe bei Patienten <math>\geq 2</math> Jahre erfasst. Bei Kindern <math>&lt; 2</math> Jahre 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 Stehhöhe bzw. Liegelänge dreimal gemessen und der Mittelwert der Messungen für die Analysen verwendet.</p> <p>Die Stehhöhe oder Liegelänge wurde als Veränderungen in cm und Veränderungen in den 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. Die Vergleichsdaten der Z-Scores basieren auf einer repräsentativen Stichprobe gesunder Kinder in den USA (= Wachstumscharts des Centers for Disease Control) und sind nicht spezifisch für Kinder mit besonderen Behandlungsanforderungen. Die Perzentile basieren ebenfalls auf den Daten aus den USA.</p> <p><b>Erhebungszeitpunkte während der kontrollierten Behandlungsphase:</b> Baseline, Wochen 24, 40 und 64 (oder Vorzeitiger Abbruch).</p>
<b>Bewertung</b>	<p><b>Operationalisierung:</b> Die Operationalisierung ist nachvollziehbar beschrieben.</p> <p><b>Validität und Patientenrelevanz:</b> Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p> <p>Die anthropometrischen Parameter können – insbesondere bei Kindern mit charakteristischen, krankheitsbedingten Wachstumsstörungen – als patientenrelevante Morbiditätsparameter eingeschätzt werden. Dabei sind Angaben, welche für das 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 Z-Scores wurden nicht berücksichtigt.</p>

<b>Endpunktkategorie Morbidität</b>	
<b>Endpunkt</b>	<b>Gehfähigkeit (6MWT)</b>
<b>Operationalisierung</b>	<p>Der 6-Minuten-Gehtest (6MWT) wurde in der Studie UX023-CL301 bei Kindern eingesetzt, die zu Baseline mindestens 5 Jahre alt waren. Laut Studienunterlagen wurde der Test grundsätzlich nach den Vorgaben der American Thoracic Society (ATS) [1] von einem geschulten Kliniker durchgeführt. Patienten wurden angewiesen, eine vorher abgemessene Strecke über 6 Minuten (wiederholt) zu laufen. Die zurückgelegte Strecke wurde nach Ablauf der 6 Minuten in Metern erfasst.</p> <p>Der pU machte im Rahmen des Stellungnahmeverfahrens weitere Angaben zu einer modifizierten Durchführung des Tests: „[Es] wurde [...] ein Anreiz für die Kinder geschaffen, indem die Kinder [...] an jeder Biegung, ohne anzuhalten, Soft-Ringe in Armbandgröße sammeln und fallen lassen konnten. Die Verwendung der Ringe ermutigte die Kinder, kontinuierlich für sechs Minuten zu gehen und nach Ende des Tests die Anzahl der gesammelten Ringe zu zählen“ [18].</p> <p><b>Erhebungszeitpunkte während der kontrollierten Behandlungsphase:</b> Screeningvisite (zu Übungszwecken), Baseline, Woche 24, 40, 64 sowie innerhalb von 3 Monaten bei vorzeitiger Beendigung der Studie.</p>
<b>Bewertung</b>	<p><b>Operationalisierung:</b> Die Operationalisierung ist mit Einschränkungen nachvollziehbar beschrieben. In den eingereichten Studienunterlagen finden sich keine Informationen dazu, ob die Operationalisierung nach den Empfehlungen der ATS-Richtlinie vollumfänglich durchgeführt werden konnte. Der pU reichte im Rahmen des Stellungnahmeverfahrens das klinische Handbuch (Clinical Evaluator Manual) ein, welches im Rahmen der Erhebungen in der Studie UX023-CL201 zum Einsatz kam. Laut Auskunft in der mündlichen Anhörung fand die Durchführung des 6MWT in der Studie UX023-CL301 analog zu Studie UX023-CL201 statt.</p> <p><b>Validität:</b> Der 6MWT wurde entwickelt zur Messung der funktionalen körperlichen Fähigkeiten bzw. körperlichen Fitness [1]. 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 Patienten mit XLH erzielt werden können. Die Reproduzierbarkeit der Testergebnisse des 6MWT ist bei qualitätssichernden Maßnahmen hoch. Zu diesen Maßnahmen gehören z. B. Schulungen des Untersuchers, standardisierte Formulierungen der Ermutigungen nach bestimmten Zeiteinheiten, Messung durch den gleichen Untersucher, ein bis zwei Probetestläufe. 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 Untersuchenden und Erfahrung der Patienten 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 XLH-Studie UX023-CL201 durchgeführt. Ergebnisse des 6MWT unterschieden sich im Median signifikant zwischen Kindern</p> <ul style="list-style-type: none"> <li>• mit einer Stehhöhe <math>\geq 2</math> SD unterhalb des Mittelwerts und <math>&lt; 2</math> SD unterhalb des Mittelwerts,</li> <li>• mit einem BOT-2-Skalenwert „Kraft und Gewandtheit“ von <math>\leq 40</math> und <math>&gt; 40</math> und</li> <li>• mit einem körperlichen Gesamt-Score (PHS-10) im SF-10 von <math>\leq 40</math> und <math>&gt; 40</math>.</li> </ul> <p>Kongruente Validität wurde ebenfalls anhand der Ergebnisse der Studie UX023-CL201 untersucht. Korrelationen des 6MWT mit dem PHS-10 und dem PSS-10 (psychischer Gesamt-Score) im SF-10, RSS-Gesamtwert und den Untertests des BOT-2 waren zu den Instrumenten und den Messzeitpunkten Baseline und Woche 40 unterschiedlich stark ausgeprägt. Die stärkste Korrelation wurde zum BOT-2-Untertest „Schnelligkeit</p>

<b>Endpunktkategorie Morbidität</b>	
	<p>und Geschicklichkeit“ zu Baseline (Pearsons <math>r = 0,858</math>), die schwächste Korrelation zum RSS-Gesamtwert zu Woche 40 (Pearsons <math>r = -0,028</math>) beobachtet [23].</p> <p>Gegenwärtig existieren Untersuchungen zur klinischen Relevanzschwelle (MID) des 6MWT in unterschiedlichen Patientenpopulationen, welche zu heterogenen Ergebnisse gelangen [4,21,22]. Untersuchungen bei XLH-Patienten liegen nicht vor. Zu diesem Zeitpunkt kann keine abschließende Einschätzung vorgenommen werden, ob eine relative Veränderung (z. B. % 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. populationsspezifisch 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 plant einen Vergleich zu Normwerten, die aus einer Untersuchung von 3- bis 11-Jährigen gesunden Kindern/Jugendlichen [9] 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 der Studie UX023-CL301 vergleichbar ist. Aufgrund des kontrollierten Studiendesigns werden die Ergebnisse im Rahmen der Studienbewertung dargestellt.</p> <p><b>Patientenrelevanz:</b> 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>
<b>Endpunkt</b>	<b>PROMIS-Domänen: Paediatric Pain Interference, Physical Function Mobility und Fatigue</b>
<b>Operationalisierung</b>	<p>Schmerzen, körperliche Funktion und Fatigue wurden in der Studie UX023-CL301 über das Patient Reported Outcomes Measurement Information System (PROMIS) erhoben. PROMIS ist ein System, das aus domänenspezifischen Instrumenten zur Beurteilung des Wohlbefindens der Patienten 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/Jugendliche im Alter von 5–17 Jahren) umfassen generische Instrumente zur Erfassung von Funktionen, Symptomen, Verhaltensweisen und Gefühlen. Dabei stehen sowohl Instrumente zur Selbsteinschätzung (für Kinder ab 8 Jahren) als auch zur Fremdbeurteilung (Elternversion bei Kindern zwischen 5 und 7 Jahren) zur Verfügung. Laut Manual besteht neben der individuellen, dynamischen Auswahl der Items aus den Itembanken (im Rahmen der Computer-adaptiven Tests (CAT)) die Möglichkeit, statische Kurzformen zu verwenden. Bei Verwendung von statischen Kurzformen beantworten die Patienten alle vorgegebenen Items einer Fragebogenversion. Die PROMIS-Entwickler schlagen für jede Itembank eine Short-Form vor. Eine eigene Itemauswahl zur Generierung einer benutzerdefinierten Kurzform ist ebenfalls möglich. Bei Verwendung des CAT-Systems werden die Items basierend auf den Antworten des Probanden computergestützt aus den Itembanken für jeden Patienten individuell zusammengestellt. Dadurch können beim CAT die Items je nach Studienteilnehmer variieren, die Scores sollen jedoch zwischen den Studienteilnehmern vergleichbar sein [27].</p> <p>In der Studie UX023-CL301 wurden Items aus den PROMIS-Itembanken der Domänen Schmerz, körperlich Funktion und Fatigue (Version 2.0) ausgewählt und eine statische Fragebogenversion zur Selbst- und Fremdeinschätzung generiert. Dabei entsprechen</p>

<b>Endpunktkategorie Morbidität</b>	
	<p>die Items nicht vollständig den Items der von den PROMIS-Entwicklern vorgeschlagenen Short-Form mit 8 bis 10 Items.</p> <ul style="list-style-type: none"> <li>• Paediatric Pain Interference: <ul style="list-style-type: none"> <li>○ 4 Items zu Problemen beim Schlafen, Rennen, Gehen, Stehen aufgrund von Schmerzen</li> <li>○ 5-Punkte-Likert-Skala (0–4) je Item: nie, fast nie, manchmal, oft, fast immer</li> </ul> </li> <li>• Physical Function Mobility: <ul style="list-style-type: none"> <li>○ 10 Items zu körperlichen Aktivitäten, Alltagsbewegungen</li> <li>○ 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</li> </ul> </li> <li>• Fatigue (Tired): <ul style="list-style-type: none"> <li>○ 8 Items zu Einschränkungen in sozialen, sportlichen und alltäglichen Aktivitäten aufgrund von Müdigkeit und Erschöpfung</li> <li>○ 5-Punkte-Likert-Skala (0–4) je Item: nie, fast nie, manchmal, oft, fast immer</li> </ul> </li> </ul> <p>Der Bezugszeitraum ist jeweils die letzte Woche.</p> <p>Der Fragebogen wurde in der Studie bei Kinder ab 5 Jahren eingesetzt. Bei Kindern, die beim Screening zwischen 5 und 7 Jahren alt waren, erfolgte die Beurteilung zu allen Erhebungszeitpunkten durch die Eltern. Bei Kindern <math>\geq 8</math> Jahre zum Zeitpunkt des Screenings wurde die Version zur Selbsteinschätzung verwendet. Die Items der Kinder- und Elternversion unterschieden sich nur in ihrer Wortwahl, nicht in der Bedeutung.</p> <p><u>Scoring und Interpretation der Werte</u></p> <p>Die Domänen-Scores wurden in der Studie als T-Werte dargestellt. Der pU hat keine weiteren Informationen zum Scoring und der Interpretation bereitgestellt. Die nachfolgenden Angaben sind dem PROMIS-Manual entnommen:</p> <p>Aus der Summe aller Itemscores einer Domäne ergeben sich die rohen Domänen-Werte. Der T-Score skaliert den Domänen-Rohwert 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:</p> <ul style="list-style-type: none"> <li>• 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.</li> <li>• Für positiv formulierte Konzepte wie bei Physical Function 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.</li> </ul> <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>Um Aussagen zur klinischen Relevanz einer Veränderung in den Domänen-Scores ableiten zu können, legt der pU Berechnungen von Hedges' g vor.</p> <p>Informationen zur psychometrischen Güte der PROMIS-Itembänke legte der pU nicht vor. Allerdings begründet der pU die Geeignetheit der aus den PROMIS-Itembänken ausgewählten Items für pädiatrische XLH anhand des Vergleichs der ausgewählten Items mit den in der Literatur beschriebenen Symptomen und Auswirkungen der XLH und den Ergebnissen qualitativer Interviews mit XLH-Patienten 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 Kinder und Jugendlichen sind und diese Aspekte mit den PROMIS-Items abgedeckt werden.</p> <p><b>Erhebungszeitpunkte während der kontrollierten Behandlungsphase:</b> Baseline, Woche 24, 40 und 64 (oder Vorzeitiger Abbruch).</p>

<b>Endpunktkategorie Morbidität</b>	
<b>Bewertung</b>	<p><b>Operationalisierung:</b> Die Operationalisierung ist weitestgehend nachvollziehbar.</p> <ul style="list-style-type: none"> <li>• Unklarheiten bestehen bezüglich der PROMIS-Domäne Pain Intensity, einer 10-Punkte-Rating-Skala zur Messung der Schmerzintensität. Laut dem COA-Dossier [23] enthielt die eingesetzte Fragebogen-Version für Kinder ab 8 Jahren diese Skala. Allerdings sind in Protokoll und Studienbericht der Studie UX023-CL301 keine Angaben zur Messung der Schmerzintensität anhand der PROMIS-Domäne Pain Intensity zu finden. Vielmehr wird in diesen Studienunterlagen angegeben, dass die Schmerzintensität mittels FPS-R erhoben wurde (s. u.).</li> <li>• Auf welche Referenzpopulation sich die T-Werte beziehen, konnte anhand der Angaben im Manual nicht eindeutig nachvollzogen werden.</li> </ul> <p><b>Validität:</b> Folgende Informationen zur Entwicklung und Validität der pädiatrischen PROMIS-Itembänke konnten anhand einer orientierenden Literaturrecherche identifiziert 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) [16].</p> <p>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=4129) mittels Item-Response-Therapie (IRT) getestet [15].</p> <p>Zur Untersuchung wurden die 293 Items randomisiert auf mehrere Test-Fragebögen (mit zum Teil überlappenden Items) aufgeteilt, um die Anzahl der von jedem Probanden zu beantwortenden Fragen 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 [7,19,28]. 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 fünf generische Gesundheitsdomänen und einer Asthma-Domäne. Davon waren 23 Items der Physical-Function-Mobility-, 23 Items der Fatigue-Tired- und 13 Items der Pain-Interference-Itembank zugeordnet [7,19,28].</p> <p>Inzwischen liegt PROMIS in der Version 2.0 vor, bei der einige Itembänke um Items ergänzt wurden. Die aktuelle Pediatric-Pain-Interference-Itembank enthält 20 Items, die Fatigue-Itembank 25 Items und die Physical-Function-Mobility 24 Items [11]. Die in der Studie verwendeten Items waren bereits in der PROMIS-Version 1.0 enthalten.</p> <p>Untersuchungen zur Konstruktvalidität und Änderungssensitivität konnten für folgende Short-Form-Skalen identifiziert werden: Pediatric Pain Interference bestehend aus 8 Items, Physical Function-Mobility mit 8 Items und Fatigue mit 10 Items [17]. Zu beachten ist, dass alle 4 in der Studie UX023-CL301 eingesetzten Pain-Intensity-Items in der Short-Form Pediatric Pain Interference enthalten sind, während nur 5 der 8 verwendeten Fatigue-Items und 5 der 10 Mobility-Items Bestandteil der entsprechenden Short-Form sind.</p> <p>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-Short-Form zum Functional Disability Inventory – Child Version und der Physical-Function-Mobility-Short-Form zur Physical-Functioning-Skala des PedsQL untersucht. In diesem laut Autoren ersten Validierungsansatz ergaben sich ähnliche Kurvenverläufe, jedoch nicht durchgehend vergleichbare Kurvenstei-</p>

<b>Endpunktkategorie Morbidität</b>	
	<p>gungen zu den Referenzinstrumenten [17]. Hinweise auf vorhandene Änderungssensitivität der PROMIS-Short-Form-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) [17]. Untersuchungen der Änderungssensitivität anhand eines patientenberichteten Ankers liegen nicht vor.</p> <p><u>Instrumente zur Fremdeinschätzung (Elternfragebogen/Proxy-Versionen)</u></p> <p>Basierend auf dem initialen pädiatrischen Itempool aus 293 Items wurden Proxy-Versionen der pädiatrischen PROMIS-Itembänke entwickelt, die eine Fremdeinschätzung der Patienten 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 Kindern 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 <math>\geq 8</math> Jahre alt; diese Kinder beantworteten parallel die Items der Selbstaussfüllerversion [14].</p> <p>Zur Untersuchung der Proxy-Itembänke wurden zunächst die Daten der Eltern von Kindern ab 8 Jahren sowie die Daten der Selbsteinschätzung dieser Kinder verwendet [29]. Konfirmatorische Faktorenanalysen und IRT-Analysen zur Skalendimensionalität 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 Short-Forms für die einzelnen Itembänke der Elternversionen entwickelt, wobei im Ergebnis die Proxy-Short-Form-Versionen jeweils ein Teilset der Kinderversionen darstellten.</p> <p>Zweidimensionale IRT-Modelle wurden eingesetzt, um die Korrelation zwischen Eltern- und Kinderversion zu schätzen. Es zeigte sich eine schwache bis moderate Übereinstimmung zwischen der Fremd- und Selbsteinschätzung (Korrelationen zwischen 0,4–0,7). Diese Ergebnisse werden von den Autoren 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 [29]. Angaben zur Änderungssensitivität der Proxy-Itembänke wurden über die orientierende Kurzrecherche nicht identifiziert.</p> <p>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 [30]. Dazu wurde das DIF zwischen den Elternantworten der Altersgruppe 5–7 Jahre und der Altersgruppe 8–17 Jahre untersucht. Datengrundlage war die Stichprobe der 1.980 Eltern gesunder und chronisch kranker Kindern zwischen 5 und 17 Jahre (s. o. [14]). Items, die DIF aufwiesen, wurden einem 7-köpfigen Expertenpanel vorgelegt. Das Expertenpanel 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. [30].</p> <p><u>Zusammenfassung</u></p> <p>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.</p>

<b>Endpunktkategorie Morbidität</b>	
	<p>Die pädiatrischen PROMIS-Itembänke werden in der Literatur als geeignete Instrumente zur Erhebung der Endpunkte Schmerzen, Fatigue und Einschränkungen in der körperlichen Funktion beschrieben. Eine abschließende Einschätzung der Validität der PROMIS-Itembänke kann an dieser Stelle nicht erfolgen. Auf folgende Aspekte wird hingewiesen: Hinweise auf eine vorhandene Änderungssensitivität der drei pädiatrischen Short-Form-Skalen Pediatric Pain Interference, Physical Function-Mobility und Fatigue liegen vor. Es sind jedoch nicht alle in der Studie abgefragten Items in den Short-Forms enthalten. Inwieweit Ergebnisse der Short-Forms auf alle Items in den drei Itembänken zur Selbsteinschätzung (und damit auch auf die in der Studie verwendeten Items und den daraus gebildeten Summenscores) übertragbar sind, kann nicht beurteilt werden. Ebenso kann keine Aussage zur Änderungssensitivität der Proxy-Items getroffen werden.</p> <p><b>Patientenrelevanz:</b> Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der <u>Verbesserung des Gesundheitszustandes</u>, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität.“</p> <p>Die 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 Selbsteinschätzung und Fremdeinschätzung zu beachten. Inwiefern diese Darstellung in der PROMIS-Methodik vorgesehen ist, ist aufgrund fehlender Angaben unklar. Grundsätzlich erscheint hierbei die unterschiedliche Einschätzung von Einschränkungen durch Eltern oder die betroffenen Kinder problematisch. Dies zeigt sich u. a. in den schwachen bis moderaten Übereinstimmungen zwischen der Fremd- und Selbsteinschätzung [29]. 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.</p>
<b>Endpunkt</b>	<b>Schmerzintensität: FPS-R</b>
<b>Operationalisierung</b>	<p>Die Faces Pain Scale – Revised (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 [13] wird das Kind vom Interviewer aufgefordert, auf ein Gesicht zu zeigen, dass am besten die Intensität des Schmerzes im aktuellen Moment abbildet (s. Abbildung 2).</p> <div style="text-align: center;">  </div> <p><i>Abbildung 2: Faces Pain Scale – Revised, bewertet 0-2-4-6-8-10 [13]</i></p> <p>Laut Angabe in Studienbericht, Studienprotokoll und SAP [25] wurde die FPS-R bei allen Kindern eingesetzt, die zu Baseline mindestens 5 Jahre alt waren. Abweichende Angaben finden sich in einem im Stimmungsverfahren eingereichten COA-Dossier [23] mit Stand vom 6. Juli 2017, laut dessen der Einsatz ausschließlich in der Altersgruppe der 5- bis 7-Jährigen stattfand.</p> <p><b>Erhebungszeitpunkte während der kontrollierten Behandlungsphase:</b> Baseline, Wochen 24, 40 und 64 (oder Vorzeitiger Abbruch).</p>

Endpunktkategorie Morbidität	
<b>Bewertung</b>	<p><b>Operationalisierung:</b> Die Operationalisierung ist mit Einschränkungen nachvollziehbar beschrieben. Durch abweichende Angaben im COA-Dossier ist unklar, ob der Fragebogen auch in der Altersgruppe der 8- bis 12-Jährigen vollumfänglich eingesetzt wurde.</p> <p><b>Validität:</b> Die Skala ist das Ergebnis einer Überarbeitung der originalen Faces Pain Scale (FPS) [2], welche Schmerzintensität auf einer 7-Punkte-Skala abbildete. Durch die Überarbeitung sollte eine Vergleichbarkeit mit 0–5 oder 0–10 Antwortformaten anderer vorhandener Schmerzskalen geschaffen werden [12].</p> <p>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 Interrater-Korrelationen zwischen 0,84–0,99 [12].</p> <p>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, 100 = sehr starke Schmerzen) eingesetzt. Es zeigten sich starke Korrelationen in der Bewertung der beiden Instrumente (<math>r = 0,93</math>; <math>p &lt; 0,001</math>). In der dritten Phase verwendeten eine Stichprobe von Kindern (im Alter von 4 bis 12 Jahren) die FPS-R, und randomisiert zusätzlich entweder</p> <ul style="list-style-type: none"> <li>• eine visuelle Analogskala (VAS) (0 = keine Schmerzen, 10 = schlimmster Schmerz) oder</li> <li>• eine farbige Analogskala (CAS) (links schmal und weiß = keine Schmerzen, rechts breit und dunkelrot = stärkste Schmerzen).</li> </ul> <p>Alle Kinder (N=90) bewerten Schmerzen während des Krankenhausaufenthaltes bei chirurgischen und nicht-chirurgischen Schmerzzuständen. Der Pearson-Korrelationskoeffizient zwischen der FPS-R und der VAS bzw. CAS lag in allen Altersgruppen über <math>r &gt; 0,7</math>. 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 die 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 den Autoren erwartet, in inverser Richtung mit der Prädiktorvariable Zeit korreliert [6].</p> <p>Der pU stellt im COA-Dossier eigene Untersuchungen zum Verständnis der FPS-R bei XLH-Erkrankten dar. Dazu wurden 4 Eltern von 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 [23].</p> <p><b>Patientenrelevanz:</b> 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. Gleichzeitig ist anzumerken, dass die Erhebung akuter Schmerzen zu Studienvisiten in einem Anwendungsgebiet mit chronischem Schmerzbild nicht das Spektrum der auftretenden Symptomatik vollumfänglich erfasst.</p>

Tabelle 4: Beschreibung der Endpunkte der Kategorie Sicherheit

Endpunktkategorie Sicherheit	
Endpunkt	Unerwünschte Ereignisse
<b>Operationalisierung</b>	<p>Ein unerwünschtes Ereignis (UE) wurde definiert als jedes unerwünschte medizinische Ereignis, das mit dem Gebrauch des Medikaments in Verbindung steht. Eine UE kann daher jedes nachteilige und unerwünschte Symptom (z. B. auch ein abnormaler Laborbefund) oder jede Erkrankung sein, die zeitlich mit der Verwendung des Arzneimittels verbunden ist, unabhängig davon, ob es mit dem Arzneimittel in Zusammenhang steht oder nicht.</p> <p>UE wurden bei allen Visiten und im gesamten Studienverlauf gemäß MedDRA (Version 18.1) erhoben. Laut originalem Studienprotokoll und SAP 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 Patienten 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 weiterhin entweder nach einem anderen Protokoll oder je nach Auflagen der Zulassung durch das Monitoring der Sicherheit gemeldet.</p> <p>Sofern angemessen, wurden klinisch signifikante Veränderungen gegenüber Baseline-Werten bei Untersuchungen, Vitalparametern, klinischen Laborparametern, Nierenultraschall, GFR, ECHO und EKG als UE erfasst.</p> <p><u>Schwerwiegende unerwünschte Ereignisse (SUE)</u> Ein SUE ist ein UE, das nach Einschätzung des Studienarztes oder des pU folgende Auswirkungen hatte:</p> <ul style="list-style-type: none"> <li>• Tod</li> <li>• lebensbedrohliches UE</li> <li>• stationärer Krankenhausaufenthalt oder Verlängerung des bestehenden Krankenhausaufenthaltes</li> <li>• anhaltende oder erhebliche Unfähigkeit oder erhebliche Beeinträchtigung der Fähigkeit normale Lebensfunktionen auszuführen</li> <li>• angeborene Anomalie/Geburtsstörung</li> </ul> <p>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.</p> <p>Medizinische Ereignisse, die nicht zum Tod führten, nicht sofort lebensbedrohlich waren oder einen Krankenhausaufenthalt erforderten, konnten zudem als schwerwiegend angesehen werden, wenn sie einen Studienteilnehmer gefährden oder einen medizinischen oder chirurgischen Eingriff erforderten, um eines der oben genannten Ereignisse zu verhindern.</p> <p><u>Schweregrad von UE</u> Soweit möglich, wurde der Schweregrad der UE anhand der Common Terminology Criteria for Adverse Events (CTCAE) des National Cancer Institute bewertet.</p> <p>Wenn ein UE nicht nach CTCAE bewertet werden konnte, sollte es als mild, mäßig, schwer, lebensbedrohlich oder Tod unter Verwendung der folgenden Definitionen eingestuft werden:</p> <ul style="list-style-type: none"> <li>• 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.</li> <li>• 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.</li> </ul>

<b>Endpunktkategorie Sicherheit</b>	
	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Lebensbedrohlich (Grad 4): Ereignisse, die den Studienteilnehmer einer unmittelbaren Todesgefahr aussetzen oder handlungsunfähig machen.</li> <li>• Tod (Klasse 5): Ereignisse, die zum Tode führen.</li> </ul> <p><u>UE von besonderem Interesse</u></p> <p>Im SAP vordefinierte Ereignisse von besonderem Interesse waren:</p> <ul style="list-style-type: none"> <li>• Reaktionen an der Injektionsstelle: definiert durch Preferred Terms im MedDRA-Oberbegriff (high-level term, HLT) "Reaktion an der Injektionsstelle".</li> <li>• Hypersensibilität: definiert durch relevante Preferred Terms im standardisierten MedDRA-Query (SMQ) für "Hypersensibilität".</li> <li>• Hyperphosphatämie: definiert durch Preferred Terms "Hyperphosphatämie" und "Blutphosphat erhöht".</li> <li>• Ektopische Mineralisierung: Da kein SMQ laut pU verfügbar ist, wurde das UE mit Hilfe einer MedDRA-Suche von „Mineralisierung“ definiert.</li> <li>• Restless-Legs-Syndrom: definiert durch die Preferred Terms "Restless-Legs-Syndrom", "Restlessness", "Akathisia", "Sinnesstörungen", "Psychomotorische Hyperaktivität", "Gliederschmerzen", "Neuromuskuläre Schmerzen" und "Formation".</li> </ul> <p><b>Erhebungszeitpunkte während der kontrollierten Behandlungsphase:</b></p> <p>Jede Studienvisite: Woche 1 und 4, danach alle 8 Wochen, weitere Hausvisiten waren je nach Nähe zum Studienzentrum und Patientenpräferenz möglich (nicht in Japan und Südkorea).</p>
<b>Bewertung</b>	<p><b>Operationalisierung:</b> Die Operationalisierung ist nachvollziehbar.</p> <p><b>Validität:</b> Der Endpunkt wird als valide einschätzt.</p> <p><b>Patientenrelevanz:</b> Entsprechend § 2 Satz 3AM -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.</p> <p>Kritisch anzumerken ist, dass mit Protokollamendment 1 die 12-wöchige Nachbeobachtung auf Patienten eingeschränkt wurde, die nach Studienende nicht anderweitig mit Burosumab therapiert werden. Da es sich bei Burosumab um eine kontinuierliche Therapie handelt, ist davon auszugehen, dass für nahezu alle Patienten keine Informationen aus der Nachbeobachtungsphase in die Auswertung der Sicherheitsendpunkte der Studie UX023-CL301 einfließen werden.</p>

### 3.4. Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 5: Verzerrungspotential auf Studienebene

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung Patient	Verblindung Behandler	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
UX023-CL301	Ja <sup>1),2)</sup>	Ja <sup>3)</sup>	Nein <sup>4)</sup>	Nein <sup>4)</sup>	Nein	Ja <sup>5),6)</sup>	Hoch

<sup>1)</sup> Randomisierungssequenz erstellt durch einen unabhängigen Drittanbieter.

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

<sup>3)</sup> Zuteilung der Randomisierung erfolgte über ein Interactive Web Response System.

<sup>4)</sup> Offenes Studiendesign.

<sup>5)</sup> Es wurde laut Aussage des pU ein Interim-Studienbericht ohne interne Qualitätskontrolle eingereicht.

<sup>6)</sup> 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.

### Verzerrungspotential auf Endpunktebene

Ausgehend vom hohen Verzerrungspotential auf Studienebene wird auch in allen Endpunkten von einem hohen Verzerrungspotential ausgegangen. Weitere Verzerrungen ergeben sich in allen patienten- und arztberichteten Endpunkten aufgrund des offenen Studiendesigns. Hinsichtlich der Endpunkte Rachitissymptomatik mittels RGI-C, der verblindet erhoben wurde, und Körpergröße, die objektiv messbar ist, erscheint das Verzerrungspotential aufgrund des offenen Studiendesigns grundsätzlich geringer, wobei ein möglicher Einfluss abschließend nicht beurteilbar ist. Problematisch erscheint zudem die Unsicherheit bezüglich der Änderungen in der Studiendurchführung, welche in Protokollamendment 1 beschrieben sind.

## 3.5. Ergebnisse

### 3.5.1. Charakterisierung der Studienpopulation

Tabelle 6: Allgemeine Angaben zum Patientenfluss der Studie UX023-CL301

UX023-CL301	Burosumab n (%)	Kontrolle n (%)
Randomisierte Patienten	29 (100)	32 (100)
Davon Patienten $\geq 5$ Jahre	15 (51,7)	20 (62,5)
Studienwoche 40 abgeschlossen	29 (100)	32 (100)
Sicherheitspopulation	29 (100)	32 (100)

Tabelle 7: Charakterisierung der Studienpopulation von UX023-CL301 zu Baseline

<b>UX023-CL301</b>	<b>Burosumab (N=29)</b>	<b>Kontrolle (N=32)</b>
<i>Alter (Jahre)<sup>1)</sup></i> MW (SD) Median (min; max)	5,8 (3,4) 5,8 (0,8; 12,8)	6,3 (3,2) 6,2 (1,2; 11,9)
<i>Geschlecht, n (%)</i> männlich weiblich	13 (44,8) 16 (55,2)	14 (43,8) 18 (56,3)
<i>Region, (%)<sup>1)</sup></i> Japan Rest der Welt	2 (6,9) 27 (93,1)	3 (9,4) 29 (90,6)
<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/m<sup>2</sup>)</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 (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 (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<sup>1)</sup></i> MW (SD) Median (min; max)	3,2 (1,0) 3,0 (2,0; 6,5)	3,2 (1,1) 3,0 (2,0; 6,5)
<i>Nierenultraschall-Wert (Skala von 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>Serum-Phosphat 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-Status, 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)
<b>Medizinische Anamnese</b>		
<i>Patienten mit jeglichem Befund<sup>2),3)</sup>, n (%)</i> Muskel- und Bindegewebserkrankungen Gastrointestinale Störungen Angeborene, familiäre und genetische Störungen Allgemeine Erkrankungen u. Beschwerden am Verabreichungsort Infektionen und Befall Erkrankungen der Atemwege, des Brustraums und des Mediastinums Chirurgische und medizinische Prozeduren	25 (86,2) 17 (58,6) 11 (37,9) 10 (34,5) 9 (31,0) 7 (24,1) 9 (31,0) 10 (34,5)	26 (81,3) 15 (46,9) 10 (31,3) 9 (28,1) 8 (25,0) 10 (31,3) 7 (21,9) 6 (18,8)

<b>UX023-CL301</b>	<b>Burosumab (N=29)</b>	<b>Kontrolle (N=32)</b>
<i>XLH-bezogene Befunde, n (%)</i>		
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)
<b>Vortherapien</b>		
<i>Patienten mit konventioneller XLH-Therapie<sup>4)</sup> vor Studienbeginn, n (%)</i>		
Ja	29 (100)	32 (100)
Nein	0	0
<i>Dauer der konventionellen XLH-Therapie<sup>4)</sup> (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-Therapie<sup>4)</sup> (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)

<sup>1)</sup> Stratifizierungsmerkmal.

<sup>2)</sup> Keine Angabe zum Bezugszeitraum im Studienbericht.

<sup>3)</sup> Dargestellt werden Systemorganklassen (MedDRA-Version 18.1) > 30 % in einem Studienarm.

<sup>4)</sup> Konventionelle Therapie umfasste Alfacalcidol, Calcitriol, Calcifediol, K-Phos Neutral, Neutra-Phos(-K), Phos-Nak, Phosphoneurol, Phosphor, Polyfusor Phosphat, Natriumphosphat, zweibasiches Natriumphosphat.

Abkürzungen: BMI: Body-Mass-Index; 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.

**Tabelle 8: Exposition mit der Studienmedikation bis zum Datenschnitt am 12. Februar 2018**

<b>UX023-CL301</b>	<b>Burosumab (N=29)</b>	<b>Kontrolle (N=32)</b>
<i>Dauer der Exposition mit der Studienmedikation (Monate)</i>		
MW (SD)	12,0 (2,0)	11,6 (2,1)
Median (min; max)	11,5 (9,5; 15,8)	11,0 (9,2; 15,7)
<i>Verabreichte Burosumab-Dosen<sup>1),2)</sup>, n (%)</i>		
20 Dosen	1 (3,4)	n.z. <sup>1)</sup>
21 Dosen	3 (10,3)	
> 21 Dosen	25 (86,2)	
<i>Burosumab-Dosis bis zum Datenschnitt, n (%)</i>		
Beibehaltung der Initialdosis von 0,8 mg/kg	23 (79,3)	k.A. <sup>1)</sup>
Dosissteigerung auf 1,2 mg/kg	6 (20,7)	
<i>Durchschnittliche tägliche Dosis an oralem Phosphat von Baseline zu Woche 40 (Elementares Phosphor mg/kg)</i>		
MW (SD)	n.z.	37,8 (17,9)
Median (min; max)		34,6 (9,0; 82,5)

<b>UX023-CL301</b>	<b>Burosumab (N=29)</b>	<b>Kontrolle (N=32)</b>
<i>Tage mit fehlender Phosphat- / aktiver Vitamin-D-Behandlung, n (%)</i>		
Phosphat		
0		26 (81,3)
1		1 (3,1)
2		3 (9,4)
5	n.z.	2 (6,3)
Vitamin D		
0		27 (84,4)
1		1 (3,1)
2		3 (9,4)
6		1 (3,1)

<sup>1)</sup> Während der kontrollierten Behandlungsperiode erhielt kein Patient der Kontrollgruppe Burosumab. Im Rahmen der anschließenden Extensionsperiode können auch Patienten der Kontrollgruppe mit Burosumab behandelt werden. Zum aktuellen Datenschnitt wird bisher ein Patient der Kontrollgruppe in der Extensionsperiode mit Burosumab behandelt.

<sup>2)</sup> 21. und darauffolgende Dosen = Dosen nach der 40. Studienwoche.

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

### Begleitmedikation

*Tabelle 9: Begleitmedikationen zum Datenschnitt am 12. Februar 2018 bei > 30 % der Patienten in einer der Behandlungsgruppen der Studie UX023-CL301*

<b>Arzneimittelgruppe<sup>1)</sup></b> Arzneimittel	<b>Burosumab (N=29)</b>	<b>Kontrolle (N=32)</b>
<b>Entzündungshemmende und antirheumatische Produkte</b>	17 (58,6)	17 (53,1)
Ibuprofen	16 (55,2)	16 (50,0)
<b>Analgetika</b>	16 (55,2)	17 (53,1)
Paracetamol	15 (51,7)	15 (46,9)
<b>Systemische Antibiotika</b>	14 (48,3)	8 (25,0)
Amoxicillin	10 (34,5)	4 (12,5)
<b>Vitamine</b>	12 (41,4)	7 (21,9)
<b>Systemische Antihistaminika</b>	12 (41,4)	6 (18,8)
<b>Husten- und Erkältungspräparate</b>	10 (34,5)	3 (9,4)

<sup>1)</sup> Entsprechend der Kategorisierungen des WHODRUG 2015Q3.

### 3.5.2. Mortalität

Todesfälle wurden im Rahmen der Erhebung von Sicherheitsendpunkten erfasst. In der Studie UX023-CL301 traten keine Todesfälle auf.

### 3.5.3. Morbidität

#### Rachitissymptomatik mittels RGI-C

Tabelle 10: Primärer Endpunkt: RGI-C-Gesamtwert zu Woche 40 in Studie UX023-CL301

UX023-CL301	Burosumab (N=29)	Kontrolle (N=32)
RGI-C-Gesamtwert zu Baseline MW (SD)	n.a. <sup>1)</sup>	n.a. <sup>1)</sup>
Veränderung im RGI-C-Gesamtwert zu Woche 40 LS Mean (SE) <sup>2)</sup>	1,9 (0,1)	0,8 (0,1)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>2)</sup>	1,1 [0,8;1,2]; <0,001	

<sup>1)</sup> Angabe nicht möglich, da der RGI-C das Ausmaß der Veränderung zu Baseline erhebt.

<sup>2)</sup> ANCOVA-Modell mit RGI-C als abhängige Variable, Behandlung und Baseline-Alter-Stratifizierungsfaktor als unabhängige Variablen und Baseline-RSS als kontinuierliche Kovariate

Abkürzungen: ANCOVA: Kovarianzanalyse; KI: Konfidenzintervall; LS: Least Squares; n.a.: nicht anwendbar; MW: Mittelwert; RGI-C: Radiographic Global Impression of Change; RSS: Rickets Severity Scale; SD: Standardabweichung; SE: Standardfehler.

#### Anthropometrische Parameter (Körpergröße)

Tabelle 11: Veränderung in der Stehhöhe bzw. Liegelänge zu Woche 40 in Studie UX023-CL301

UX023-CL301	Burosumab <sup>1)</sup> (N=29)	Kontrolle (N=32)
Z-Score zu Baseline MW (SD)	(N=28) -2,3 (1,8)	-2,1 (0,9)
Veränderung im Z-Score zu Woche 40 LS Mean (SE) <sup>2)</sup>	(N=28) 0,2 (0,1)	0,1 (0,1)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>2)</sup>	0,1 [-0,1;0,2]; 0,3507	

<sup>1)</sup> Für einen Patienten im Burosumab-Arm war keine Angabe zu Baseline verfügbar. Laut SAP wurden fehlende Werte als solche behandelt. Der Umgang mit fehlenden Werten im GEE-Modell bleibt damit unklar. Inwiefern die Bedingung „Missing Completely at Random“ erfüllt ist, ist unklar.

<sup>2)</sup> 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. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.

Abkürzungen: GEE: Generalized Estimation Equation; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; RSS: Rickets Severity Scale; SAP: statistischer Analyseplan; SD: Standardabweichung; SE: Standardfehler.

## Gefähigkeit (6MWT)

Tabelle 12: Veränderung der Laufstrecke im 6MWT zu Woche 40 in Studie UX023-CL301, Patienten  $\geq 5$  Jahre der ITT-Population

UX023-CL301	Burosumab (N=15 <sup>1)</sup> )	Kontrolle (N=20 <sup>1)</sup> )
6MWT-Distanz zu Baseline (Meter) MW (SD)	365,9 (118,1)	450,5 (106,4)
Veränderung der 6MWT-Distanz zu Woche 40 (Meter) LS Mean (SE) <sup>2)</sup>	52,1 (16,2)	10,6 (14,1)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>2),3)</sup>	41,5 [-0,6;83,6]; 0,0534	
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 40 LS Mean (SE) <sup>2)</sup>	6,2 (2,7)	0,5 (2,4)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>2),3)</sup>	5,6 [-1,2;12,4]; 0,1058	

<sup>1)</sup> Der 6MWT wurde ausschließlich bei Patienten durchgeführt, die zum Zeitpunkt der Screening-Visite mindestens 5 Jahre alt waren.

<sup>2)</sup> GEE-Modell mit Veränderung von Baseline als abhängige Variable, Studienvisite, Behandlungsarm, Interaktion zwischen Behandlungsgruppe und Studienvisite und Baseline-RSS als Faktoren und 6MWT bei Baseline als Kovariable, mit austauschbarer Kovarianzstruktur. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.

<sup>3)</sup> Laut Studienbericht fehlten für zwei Patienten des Burosumab-Studienarms Daten: Ein Patient absolvierte den 6MWT nicht zu Baseline, ein Patient nicht zu Woche 40. Laut SAP wurden fehlende Werte als solche behandelt. Der Umgang mit fehlenden Werten im GEE-Modell bleibt damit unklar. Inwiefern die Bedingung „Missing Completely at Random“ erfüllt ist, ist unklar. Eine nicht a priori vorgesehene Analyse mit Ausschluss der beiden Patienten mit fehlenden Daten zeigte ebenfalls statistisch nicht signifikante Ergebnisse der gleichen Richtung und vergleichbarem Ausmaßes.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; 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.

## PROMIS-Domänen: Schmerzen, körperliche Funktion und Fatigue

Tabelle 13: PROMIS-Scores zu Woche 40 in Studie UX023-CL301, Patienten  $\geq 5$  Jahre der ITT-Population

UX023-CL301	Burosumab (N=15 <sup>1)</sup> )	Kontrolle (N=20 <sup>1)</sup> )
<b>Paediatric Pain Interference</b>		
Baseline T-Score <sup>2),3)</sup> MW (SD)	53,1 (11,0)	49,9 (12,1)
Veränderung im T-Score zu Woche 40 LS Mean (SE) <sup>4)</sup>	-5,3 (1,7)	-0,3 (1,5)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>4)</sup> Hedges' g [95%-KI]; p-Wert <sup>5)</sup>	-5,0 [-9,3;-0,8]; 0,0212 -0,7 [-1,4;0,0]; 0,0404	
<b>Physical Function Mobility</b>		
Baseline T-Score <sup>3),6)</sup> MW (SD)	45,2 (9,1)	45,5 (9,9)
Veränderung im T-Score zu Woche 40 LS Mean (SE) <sup>4)</sup>	2,8 (1,3)	0,1 (1,0)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>4)</sup>	2,7 [-0,5;5,9]; 0,1009	

UX023-CL301	Burosumab (N=15 <sup>1)</sup> )	Kontrolle (N=20 <sup>1)</sup> )
<b>Fatigue</b>		
<i>Baseline T-Score</i> <sup>3),7)</sup> MW (SD)	48,8 (9,6)	47,0 (13,7)
<i>Veränderung im T-Score zu Woche 40</i> LS Mean (SE) <sup>4)</sup>	-4,3 (1,7)	-1,05 (1,8)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>4)</sup>	-3,3 [-7,9;1,4]; 0,1676	

<sup>1)</sup> Der PROMIS-Fragebogen wurde ausschließlich bei Patienten eingesetzt, die zum Zeitpunkt der Screening-Visite mindestens 5 Jahre alt waren.

<sup>2)</sup> Höhere Werte sind gleichbedeutend mit einer stärkeren Beeinträchtigung durch Schmerzen.

<sup>3)</sup> T-Score = standardisierter Score mit Mittelwert von 50 und Standardabweichung von 10.

<sup>4)</sup> GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und Baseline-PROMIS-Domänen-Score als kontinuierliche Kovariate. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.

<sup>5)</sup> Angabe entnommen aus Stellungnahme des pU [18].

<sup>6)</sup> Höhere Werte sind gleichbedeutend mit einer besseren körperlichen Aktivität.

<sup>7)</sup> Höhere Werte sind gleichbedeutend mit einer stärkeren Fatigue.

Abkürzungen: GEE: Generalized Estimation Equation; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; RSS: Rickets Severity Scale; SAP: statistischer Analyseplan; SD: Standardabweichung; SE: Standardfehler.

### Schmerzintensität (FPS-R)

Tabelle 14: Veränderung der akuten Schmerzintensität mittels FPS-R zu Woche 40 in Studie UX023-CL301, Patienten  $\geq 5$  Jahre der ITT-Population

UX023-CL301	Burosumab (N=15 <sup>1)</sup> )	Kontrolle (N=20 <sup>1)</sup> )
<i>FPS-R-Wert zu Baseline</i> <sup>2)</sup> MW (SD)	0,4 (1,1)	0,7 (1,2)
<i>Veränderung des FPS-R-Werts zu Woche 40</i> LS Mean (SE) <sup>3)</sup>	0,0 (0,3)	0,0 (0,3)
LS-Mean-Differenz [95%-KI]; p-Wert <sup>3)</sup>	0,0 [-0,8;0,8]; 0,9862	

<sup>1)</sup> Die FPS-R wurde ausschließlich bei Patienten eingesetzt, die zum Zeitpunkt der Screening-Visite mindestens 5 Jahre alt waren.

<sup>2)</sup> Höhere Werte sind gleichbedeutend mit einer stärkeren Schmerzintensität; Skala reicht von 0 bis 10.

<sup>3)</sup> 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 Kovariable, mit austauschbarer Kovarianzstruktur. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.

Abkürzungen: 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.

### 3.5.4. Lebensqualität

In der Studie UX023-CL301 wurde Lebensqualität durch den SF-10 erhoben. Aufgrund von Einschränkungen in den psychometrischen Eigenschaften werden die Ergebnisse der Fragebogenerhebung nicht dargestellt [10].

### 3.5.5. Sicherheit

Tabelle 15: Zusammenfassung der UE zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301

Patienten mit mindestens einem ...	Burosumab (N=29) n (%)	Kontrolle <sup>1)</sup> (N=32) n (%)	Relatives Risiko [95%-KI]
UE	28 (96,6)	26 (81,3)	.. <sup>2)</sup>
UE CTCAE-Grad $\geq$ 3	4 (13,8)	2 (6,3)	2,2 [0,4;11,2] <sup>3)</sup>
SUE	3 (10,3)	1 (3,1)	3,3 [0,4;30,1] <sup>3)</sup>
UE, das zum Abbruch der Studienmedikation führte	0	0	-
UE, das zum Tod führte	0	0	-

<sup>1)</sup> Keine Angaben zur Auswertung des geplanten Cross-Over vom Kontroll- in den Burosumab-Arm nach Woche 64 verfügbar. Laut Studienbericht befand sich ein Patient der Kontrollgruppe zum Zeitpunkt des Datenschnitts in Studienwoche 66 und hatte in den Burosumab-Arm gewechselt.

<sup>2)</sup> Patientenrelevanz nicht eindeutig beurteilbar.

<sup>3)</sup> Eigene Berechnungen mit RevMan 5.2. Angaben in den Unterlagen des pU nicht verfügbar.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; KI: Konfidenzintervall; pU: pharmazeutischer Unternehmer; (S)UE: (schwerwiegende/s) unerwünschte/s Ereignis/se.

Tabelle 16: UE mit Inzidenz  $\geq$  10 % in mindestens einem Studienarm zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301

MedDRA-Systemorganklasse <sup>1)</sup> Preferred Term	Burosumab (N=29) n (%)	Kontrolle <sup>2)</sup> (N=32) n (%)
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>22 (75,9)</b>	<b>8 (25,0)</b>
Pyrexie	15 (51,7)	6 (18,8)
Reaktion an der Injektionsstelle	7 (24,1)	0
Erythem der Injektionsstelle	6 (20,7)	0
Pruritus an der Injektionsstelle	3 (10,3)	0
Ausschlag an der Injektionsstelle	3 (10,3)	0
<b>Infektionen und parasitäre Erkrankungen</b>	<b>22 (75,9)</b>	<b>23 (71,9)</b>
Nasopharyngitis	10 (34,5)	13 (40,6)
Zahnabszess <sup>3)</sup>	7 (24,1)	3 (9,4)
Influenza	4 (13,8)	6 (18,8)
Rhinitis	3 (10,3)	2 (6,3)
Infektion der oberen Atemwege	3 (10,3)	2 (6,3)
<b>Gastrointestinale Störungen</b>	<b>20 (69,0)</b>	<b>15 (46,9)</b>
Übelkeit	11 (37,9)	8 (25,0)
Diarrhoea	7 (24,1)	2 (6,3)
Zahnkaries <sup>3)</sup>	5 (17,2)	2 (6,3)
Zahnschmerzen <sup>3)</sup>	4 (13,8)	1 (3,1)
Schmerzen im Oberbauch	3 (10,3)	3 (9,4)
Zahnen	3 (10,3)	0

<b>MedDRA-Systemorganklasse<sup>1)</sup></b> Preferred Term	<b>Burosumab (N=29)</b> n (%)	<b>Kontrolle<sup>2)</sup></b> (N=32) n (%)
<b>Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>	<b>19 (65,5)</b>	<b>9 (28,1)</b>
Husten	12 (41,4)	6 (18,8)
Rhinorrhoea	5 (17,2)	2 (6,3)
Asthma	4 (13,8)	1 (3,1)
Oropharyngeale Schmerzen	4 (13,8)	1 (3,1)
Nasenkongestion	3 (10,3)	1 (3,1)
<b>Erkrankungen der Muskel, des Skeletts und des Bindegewebes<sup>3)</sup></b>	<b>15 (51,7)</b>	<b>13 (40,6)</b>
Arthralgie	11 (37,9)	7 (21,9)
Schmerzen in den Extremitäten	10 (34,5)	9 (28,1)
<b>Erkrankungen des Nervensystems</b>	<b>10 (34,5)</b>	<b>8 (25,0)</b>
Kopfschmerzen	8 (27,6)	5 (15,6)
<b>Erkrankungen der Haut und des Unterhautzellgewebes</b>	<b>10 (34,5)</b>	<b>2 (6,3)</b>
Ausschlag	3 (10,3)	2 (6,3)
<b>Verletzungen, Vergiftungen und prozedurale Komplikationen</b>	<b>8 (27,6)</b>	<b>2 (6,3)</b>
Kontusion	4 (13,8)	0
<b>Untersuchungen</b>	<b>8 (27,6)</b>	<b>4 (12,5)</b>
Vitamin D vermindert	6 (20,7)	1 (3,1)
<b>Ohren- und Labyrinthstörungen</b>	<b>5 (17,2)</b>	<b>2 (6,3)</b>
Ohrenschmerzen	3 (10,3)	1 (3,1)
<b>Störungen des Immunsystems</b>	<b>5 (17,2)</b>	<b>5 (15,6)</b>
Saisonale Allergie	4 (13,8)	3 (9,4)
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>5 (17,2)</b>	<b>3 (9,4)</b>
Vitamin-D-Mangel	4 (13,8)	1 (3,1)
<b>Angeborene, familiäre und genetische Störungen<sup>3)</sup></b>	<b>3 (10,3)</b>	<b>1 (3,1)</b>

<sup>1)</sup> MedDRA-Version 18.1.

<sup>2)</sup> Keine Angaben zur Auswertung des geplanten Cross-Over vom Kontroll- in den Burosumab-Arm nach Woche 64 verfügbar. Laut Studienbericht befand sich ein Patient der Kontrollgruppe zum Zeitpunkt des Datenschnitts in Studienwoche 66 und hatte in den Burosumab-Arm gewechselt.

<sup>3)</sup> Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidität“.

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

Tabelle 17: UE mit CTCAE-Grad  $\geq 3$  zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301

MedDRA-Systemorganklasse <sup>1)</sup> Preferred Term	Burosumab (N=29) n (%)	Kontrolle <sup>2)</sup> (N=32) n (%)
<b>Infektionen und parasitäre Erkrankungen</b> Virale Gastroenteritis	1 (3,4)	0
<b>Untersuchungen</b> Ketonkörper im Urin	1 (3,4)	0
<b>Erkrankungen der Muskel, des Skeletts und des Bindegewebes<sup>3)</sup></b> Arthralgie	1 (3,4)	1 (3,1)
<b>Nieren- und Harnwegserkrankungen</b> Dysurie	1 (3,4)	0
<b>Störungen des Immunsystems</b> Nahrungsmittelallergie	0	1 (3,4)

<sup>1)</sup> MedDRA-Version 18.1.

<sup>2)</sup> Keine Angaben zur Auswertung des geplanten Cross-Over vom Kontroll- in den Burosumab-Arm nach Woche 64 verfügbar. Laut Studienbericht befand sich ein Patient der Kontrollgruppe zum Zeitpunkt des Datenschnitts in Studienwoche 66 und hatte in den Burosumab-Arm gewechselt.

<sup>3)</sup> Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidity“.

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

Tabelle 18: SUE zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301

MedDRA-Systemorganklasse <sup>1)</sup> Preferred Term	Burosumab (N=29) n (%)	Kontrolle <sup>2)</sup> (N=32) n (%)
<b>Infektionen und parasitäre Erkrankungen</b> Virale Infektion	1 (3,4)	0
<b>Erkrankungen des Nervensystems</b> Migräne	1 (3,4)	0
<b>Angeborene, familiäre und genetische Störungen<sup>3)</sup></b> Kraniosynostose	1 (3,4)	0
<b>Nieren- und Harnwegserkrankungen</b> Hämaturie	0	1 (3,1)

<sup>1)</sup> MedDRA-Version 18.1.

<sup>2)</sup> Keine Angaben zur Auswertung des geplanten Cross-Over vom Kontroll- in den Burosumab-Arm nach Woche 64 verfügbar. Laut Studienbericht befand sich ein Patient der Kontrollgruppe zum Zeitpunkt des Datenschnitts in Studienwoche 66 und hatte in den Burosumab-Arm gewechselt.

<sup>3)</sup> Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidity“.

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

*Tabelle 19: UE von besonderem Interesse zum Zeitpunkt des Datenschnitts am 12. Februar 2018 in der Studie UX023-CL301*

<b>UE von besonderem Interesse<sup>1)</sup></b>	<b>Burosumab (N=29) n (%)</b>	<b>Kontrolle<sup>2)</sup> (N=32) n (%)</b>
Reaktionen an der Injektionsstelle	13 (44,8)	k.A. <sup>3)</sup>
Hypersensibilität	9 (31,0)	3 (9,4)
Ektopische Mineralisierung	0	0
Hyperphosphatämie	0	0
Restless-Legs-Syndrom	0	0

<sup>1)</sup> Abweichungen zu Tabelle 16 ergeben sich aus Erhebung der UE von besonderem Interesse über HLT, SMQ und vom pU erstellte Kategorisierungen von Preferred Terms gemäß MedDRA (Version 18.1.) (s. Tabelle 4).

<sup>2)</sup> Keine Angaben zur Auswertung des geplanten Cross-Over vom Kontroll- in den Burosumab-Arm nach Woche 64 verfügbar. Laut Studienbericht befand sich ein Patient der Kontrollgruppe zum Zeitpunkt des Datenschnitts in Studienwoche 66 und hatte in den Burosumab-Arm gewechselt.

<sup>3)</sup> 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: HLT: High Level Term; k.A.: keine Angabe; MedDRA: Medical Dictionary for Regulatory Activities; pU: pharmazeutischer Unternehmer; SAP: statistischer Analyseplan; SMQ: Standardised MedDRA Query; UE: unerwünschte/s Ereignis/se.

## 4. Fazit

Tabelle 20: Zusammenfassende Darstellung der Ergebnisse der Studie UX023-CL301

Endpunkt	Ergebnis	Effekt
<b>Morbidität</b>		
Veränderung der Stehhöhe zwischen Baseline und Woche 40 (Z-Score) <sup>1)</sup> LS Mean (SE)	Burosumab: 0,2 (0,1) Kontrolle: 0,1 (0,1)	↔
LS Mean-Differenz [95%-KI]; p-Wert <sup>2),3)</sup>	0,1 [-0,1;0,2]; 0,3507	
Veränderung der 6MWT-Distanz zwischen Baseline und Woche 40 (Meter) <sup>4)</sup> LS Mean (SE)	Burosumab: 52,1 (16,2) Kontrolle: 10,6 (14,1)	↔
LS Mean-Differenz [95%-KI]; p-Wert <sup>5),6)</sup>	41,5 [-0,6;83,6]; 0,0534	
Veränderung des Anteils der erwarteten 6MWT-Distanz zwischen Baseline und Woche 40 (in %) <sup>4)</sup> LS Mean (SE)	Burosumab: 6,2 (2,7) Kontrolle: 0,5 (2,4)	↔
LS Mean-Differenz [95%-KI]; p-Wert <sup>5),6)</sup>	5,6 [-1,2;12,4]; 0,1058	
Veränderung der Symptome mittels PROMIS zwischen Baseline und Woche 40 <sup>4)</sup> LS-Mean-Differenz zwischen Studienarmen [95%-KI]; p-Wert <sup>7)</sup>	<u>Paediatric Pain Interference Score<sup>8)</sup></u> : -5,0 [-9,3;-0,8]; 0,0212 <u>Physical Function Mobility Score<sup>9)</sup></u> : 2,7 [-0,5;5,9]; 0,1009 <u>Fatigue Score<sup>10)</sup></u> : -3,3 [-7,9;1,4]; 0,1676	↑ <sup>11)</sup> ↔ ↔
Veränderung des FPS-R-Werts für akute Schmerzintensität <sup>4),12)</sup> LS Mean (SE)	Burosumab: 0,0 (0,3) Kontrolle: 0,0 (0,3)	↔
LS-Mean-Differenz [95%-KI]; p-Wert <sup>13)</sup>	0,0 [-0,8;0,8]; 0,9862	
<b>Sicherheit<sup>14)</sup></b>		
UE CTCAE-Grad ≥ 3 Burosumab vs. Kontrollarm, n (%) RR [95%-KI] <sup>15)</sup>	4 (13,8) vs. 2 (6,3) 2,2 [0,4;11,2]	↔
SUE Burosumab vs. Kontrollarm, n (%) RR [95%-KI] <sup>15)</sup>	3 (10,3) vs. 1 (3,1) 3,3 [0,4;30,1]	↔
UE, das zum Abbruch der Studienmedikation führte Burosumab vs. Kontrollarm, n (%)	keine Ereignisse	n.b.
UE, das zum Tod führte Burosumab vs. Kontrollarm, n (%)	keine Ereignisse	n.b.

Endpunkt	Ergebnis	Effekt
<b>UE von besonderem Interesse</b> Burosumab vs. Kontrollarm, n (%)		
Reaktionen an der Injektionsstelle	13 (44,8) vs. k.A. <sup>16)</sup>	n.b.
Hypersensibilität	9 (31,0) vs. 3 (9,4)	n.b.
Ektopische Mineralisierung	keine Ereignisse	n.b.
Hyperphosphatämie	keine Ereignisse	n.b.
Restless-Legs-Syndrom	keine Ereignisse	n.b.

- <sup>1)</sup> Auswertung in der FAS-Population mit allen randomisierten Patienten. Burosumab: N=29, Kontrolle: N=32.
- <sup>2)</sup> 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. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.
- <sup>3)</sup> Für einen Patienten im Burosumab-Arm fehlte die Angabe zu Baseline. Laut SAP wurden fehlende Werte als solche behandelt. Inwiefern die Bedingung „Missing Completely at Random“ erfüllt ist, ist unklar.
- <sup>4)</sup> Auswertung im Full Analysis Subset mit allen Patienten  $\geq 5$  Jahre zu Baseline. Burosumab: N=15, Kontrolle: N=20.
- <sup>5)</sup> GEE-Modell mit Veränderung von Baseline als abhängige Variable, Studienvisite, Behandlungsarm, Interaktion zwischen Behandlungsgruppe und Studienvisite und Baseline-RSS als Faktoren und 6MWT bei Baseline als Kovariable, mit austauschbarer Kovarianzstruktur. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.
- <sup>6)</sup> Laut Studienbericht fehlten für zwei Patienten des Burosumab-Studienarms Daten: Ein Patient absolvierte den 6MWT nicht zu Baseline, ein Patient nicht zu Woche 40. Laut SAP wurden fehlende Werte als solche behandelt. Inwiefern die Bedingung „Missing Completely at Random“ erfüllt ist, ist unklar. Eine nicht a priori vorgesehene Analyse mit Ausschluss der beiden Patienten mit fehlenden Daten zeigte ebenfalls statistisch nicht signifikante Ergebnisse der gleichen Richtung.
- <sup>7)</sup> GEE-Modell mit Behandlung, Visite, Interaktion Behandlung x Visite und Baseline-RSS-Stratifizierungsfaktor als unabhängige Variablen und Baseline-PROMIS-Domänen-Score als kontinuierliche Kovariate. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.
- <sup>8)</sup> Höhere Werte sind gleichbedeutend mit einer stärkeren Beeinträchtigung durch Schmerzen.
- <sup>9)</sup> Höhere Werte sind gleichbedeutend mit einer besseren körperlichen Aktivität.
- <sup>10)</sup> Höhere Werte sind gleichbedeutend mit einer stärkeren Fatigue.
- <sup>11)</sup> Klinische Relevanz unklar. Hedges' g: -0,7 [95%-KI: -1,4;0,0]; p=0,0404 (Angabe entnommen aus Stellungnahme des pU [17]).
- <sup>12)</sup> Höhere Werte sind gleichbedeutend mit einer stärkeren Schmerzintensität, Skala reicht von 0 bis 10.
- <sup>13)</sup> 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 Kovariable, mit austauschbarer Kovarianzstruktur. Aus dem SAP geht nicht hervor, ob Messungen zu Woche 24 im GEE-Modell berücksichtigt wurden.
- <sup>14)</sup> Auswertung bis zum Datenschnitt am 12. Februar 2018. Keine Angaben zur Auswertung des geplanten Cross-Over vom Kontroll- in den Burosumab-Arm nach Woche 64 verfügbar. Laut Studienbericht befand sich ein Patient der Kontrollgruppe zum Zeitpunkt des Datenschnitts in Studienwoche 66 und hatte in den Burosumab-Arm gewechselt.
- <sup>15)</sup> Eigene Berechnungen mit RevMan 5.2. Angaben in den Unterlagen des pU nicht verfügbar.
- <sup>16)</sup> Laut Studienbericht waren Reaktionen an der Injektionsstelle kein UE von besonderem Interesse in der Kontrollgruppe. Diese Einschränkung findet sich nicht im SAP oder im 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; k.A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; n.b.: nicht bewertbar; PROMIS: Patient-Reported Outcomes Measurement Information System; pU: pharmazeutischer Unternehmer; RR: Relatives Risiko; RSS: Ricketts Severity Scale; SAP: statistischer Analyseplan; SE: Standardfehler; (S)UE: (schwerwiegende/s) unerwünschte/s Ereignis/se.

## Referenzen

1. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002;166(1):111-117.
2. **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.
3. **Biomedical Systems.** Ultragenyx RGI-C variability assessment for study UX023-CL201: Version 1.0 [unveröffentlicht]. 2017.
4. **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.
5. **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.
6. **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.
7. **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.
8. **European Medicines Agency (EMA).** Crysvita: product information EMEA/H/C/004275; Anhang I-IV [online]. 19.02.2018. London (GBR): EMA; 2018. [Zugriff: 08.05.2018]. URL: [http://www.ema.europa.eu/docs/de\\_DE/document\\_library/EPAR\\_-\\_Product\\_Information/human/004275/WC500245537.pdf](http://www.ema.europa.eu/docs/de_DE/document_library/EPAR_-_Product_Information/human/004275/WC500245537.pdf).
9. **Geiger R, Strasak A, Tremel B, Gasser K, Kleinsasser A, Fischer V, et al.** Six-minute walk test in children and adolescents. *J Pediatr* 2007;150(4):395-399.
10. **Gemeinsamer Bundesausschusses (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: 03.09.2018]. URL: [https://www.g-ba.de/downloads/92-975-2392/2018-04-15\\_Nutzenbewertung-G-BA\\_Burosumab-D-349.pdf](https://www.g-ba.de/downloads/92-975-2392/2018-04-15_Nutzenbewertung-G-BA_Burosumab-D-349.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. **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](https://s3.amazonaws.com/rdcms-iasp/files/production/public/Content/ContentFolders/Resources2/FPSR/facepainscale_english_eng-au-ca.pdf).

14. **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.
15. **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.
16. **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.
17. **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.
18. **Kyowa Kirin.** Stellungnahme zur Nutzenbewertung nach § 35a SGB V und Kosten-Nutzen-Bewertung nach § 35b SGB V: Stellungnahme zu Burosumab/Crysvita [unveröffentlicht]. 2018.
19. **Lai JS, Stucky BD, Thissen D, Varni JW, DeWitt EM, Irwin DE, et al.** Development and psychometric properties of the PROMIS pediatric fatigue item banks. *Qual Life Res* 2013;22(9):2417-2427.
20. **Linglart A, Biosse-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.
21. **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.
22. **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.
23. **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.
24. **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.
25. **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 40 (Primary analysis) [unveröffentlicht]. 2018.

26. **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); Protocol [unveröffentlicht]. 2016.
27. **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.
28. **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.
29. **Varni JW, Thissen D, Stucky BD, Liu Y, Gorder H, Irwin DE, et al.** PROMIS Parent Proxy Report Scales: an item response theory analysis of the parent proxy report item banks. Qual Life Res 2012;21(7):1223-1240.
30. **Varni JW, Thissen D, Stucky BD, Liu Y, Magnus B, Quinn H, et al.** PROMIS 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.