

# Nutzenbewertung



**Gemeinsamer  
Bundesausschuss**

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

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

**5. Kapitel § 12 Nr. 1 Satz 2 VerfO**

**Wirkstoff: Burosumab**

Datum der Veröffentlichung: 1. Februar 2021

## Inhaltsverzeichnis

Tabellenverzeichnis.....	3
Abkürzungsverzeichnis .....	5
Hintergrund.....	7
1 Fragestellung.....	8
2 Auswahl und Methodik der relevanten Studien.....	9
2.1 Studienbasis für die Nutzenbewertung .....	9
2.2 Aufbau und Design der eingeschlossenen Studie .....	10
2.3 Endpunkte.....	16
2.3.1 Mortalität.....	17
2.3.2 Morbidität.....	18
2.3.3 Lebensqualität.....	31
2.3.4 Sicherheit .....	31
2.3.5 Übersicht der Erhebungszeitpunkte.....	33
2.4 Statistische Methoden .....	34
2.5 Verzerrungspotential auf Studien- und Endpunktebene .....	37
3 Ergebnisse der eingeschlossenen Studie.....	39
3.1 Studiencharakteristika und Studienmedikation .....	39
3.2 Mortalität.....	46
3.3 Morbidität.....	46
3.4 Lebensqualität .....	53
3.5 Sicherheit.....	53
4 Diskussion der Methodik und Ergebnisse.....	57
4.1 Zulassungsstatus und Zulassungspopulation von Burosumab.....	57
4.2 Design und Methodik der Studie CL303 .....	57
4.3 Mortalität .....	60
4.4 Morbidität .....	60
4.5 Lebensqualität .....	65
4.6 Sicherheit.....	65
5 Anforderung an eine qualitätsgesicherte Anwendung .....	68
6 Zusammenfassung der Nutzenbewertung .....	69
Referenzen .....	71
Anhang .....	75

## Tabellenverzeichnis

Tabelle 1: Übersicht über die Studienbasis .....	9
Tabelle 2: Charakterisierung der Studie CL303 .....	10
Tabelle 3: Charakterisierung der Interventionen in der Studie CL303 während der Placebo-kontrollierten Behandlungsperiode .....	13
Tabelle 4: Übersicht der Protokollamendments .....	15
Tabelle 5: Zusammenfassung der Endpunktbewertung.....	16
Tabelle 6: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie CL303.....	33
Tabelle 7: Relevante Analysepopulationen der Studie CL303.....	35
Tabelle 8: Verzerrungspotential der Studie CL303 während der kontrollierten Behandlungsperiode.....	37
Tabelle 9: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie CL303 während der kontrollierten Behandlungsperiode .....	37
Tabelle 10: Allgemeine Angaben zum Patientenfluss in der Studie CL303.....	39
Tabelle 11: Protokollverletzungen in der Studie CL303 bis zum Datenschnitt 22. Dezember 2016 ....	40
Tabelle 12: Charakterisierung der Studienpopulation der Studie CL303 .....	41
Tabelle 13: Angaben zur Exposition mit der Studienmedikation in der Studie CL303 während der kontrollierten Behandlungsperiode .....	44
Tabelle 14: Begleitmedikation > 20 % in einem Studienarm in der Studie CL303 .....	45
Tabelle 15: Personen mit mittlerem Serumphosphat Spiegel $\geq 2,5$ mg/dL in der Mitte des Dosiszyklus während der Behandlungsperiode, Full Analysis Set <sup>1)</sup> .....	46
Tabelle 16: Veränderung im 6MWT zu Woche 24, Full Analysis Set <sup>1)</sup> .....	47
Tabelle 17: Veränderung im WOMAC zu Woche 24, Full Analysis Set <sup>1)</sup> .....	48
Tabelle 18: Personen mit Response im BPI-SF Item 3 zu Woche 24, Full Analysis Set <sup>1)</sup> .....	50
Tabelle 19: Veränderung im BPI-SF zu Woche 24, Full Analysis Set <sup>1)</sup> .....	50
Tabelle 20: Veränderung im BFI <sup>1)</sup> zu Woche 24, Full Analysis Set <sup>2)</sup> .....	51
Tabelle 21: Veränderung im PGI-I zu Woche 24, Full Analysis Set <sup>1)</sup> .....	52
Tabelle 22: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsperiode im Safety Set <sup>1)</sup> .....	53
Tabelle 23: Personen mit UE während der Behandlungsperiode mit Inzidenz $\geq 10$ % in einem Studienarm, Safety Set <sup>1)</sup> .....	54
Tabelle 24: Personen mit UE von besonderem Interesse während der Behandlungsperiode, Safety Set <sup>1)</sup> .....	56

Tabelle 25: Zusammenfassende Darstellung der Ergebnisse der Studie UX023-CL303 für die Behandlungsperiode .....	69
Tabelle 26: Personen mit mittlerem Serumphosphatspiegel $\geq 2,5$ mg/dL in der Mitte des Dosiszyklus während der Behandlungsperiode inkl. Relativem Risiko, Full Analysis Set <sup>1)</sup> .....	75
Tabelle 27: Personen mit Response im BPI-SF Item 3 zu Woche 24 inkl. Relativen Risiken, Full Analysis Set <sup>1)</sup> .....	75
Tabelle 28: Personen mit UE während der Behandlungsperiode mit Inzidenz $\geq 10$ % in einem Studienarm inkl. Relativen Risiken, Safety Set <sup>1)</sup> .....	75

## Abkürzungsverzeichnis

6MWT	6-Minuten-Gehtest
ANCOVA	Kovarianzanalyse
ATC	Anatomisch-Therapeutisch-Chemisch
BFI	Brief Fatigue Inventory
BOCF	Baseline Observation Carried Forward
BPI-SF	Brief Pain Inventory-Short Form
CL303	UX023-CL303
cm	Centimeter
COA	Clinical Outcome Assessment
CTCAE	Common Terminology Criteria for Adverse Events
dL	Deziliter
EMA	European Medicines Agency
EPAR	European Public Assessment Report
FAS	Full Analysis Set
FDA	Food and Drug Administration
FI	Fachinformation
G-BA	Gemeinsamer Bundesausschuss
GEE	General Estimating Equations
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
IWRS	Interactive Web Response System
k. A.	Keine Angabe
kg	Kilogramm
KI	Konfidenzintervall
L	Liter
LLN	Lower Limit of Normal (untere Grenze des Normalwerts)
LOCF	Last Observation Carried Forward
LS	Least Squares
max	Maximum
mBOCF	Modified Baseline Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligramm
MID	Minimal important difference
min	Minimum

mmol	Millimol
MW	Mittelwert
N	Anzahl
n. v.	Nicht verfügbar
n. z.	Nicht zutreffend
NCI	National Cancer Institute
ng	Nanogramm
NRS	Numerische Rating-Skala
PGI-I	Patient Global Impression of Improvement
PHEX	Phosphate-regulating neutral endopeptidase
PRO	Patient-reported outcomes
PT	Preferred Term/s
pU	Pharmazeutischer Unternehmer
RCT	Randomized controlled trial/s (randomisierte kontrollierte Studie/n)
ROC	Receiver operating characteristic
SAP	Statistischer Analyseplan
SD	Standardabweichung
SE	Standardfehler
SGB	Sozialgesetzbuch
SOC	Systemorganklasse/n
SUE	Schwerwiegende/s unerwünschte/s Ereignis/se
TEAE	Treatment Emergent Adverse Events
TUG	Timed Up and Go-Test
UE	Unerwünschte/s Ereignis/se
ULN	Upper Limit of Normal (obere Grenze des Normalbereichs)
VAS	Visuelle Analogskala
VerfO	Verfahrensordnung
vs.	Versus
WOMAC	Western Ontario and McMaster Universities Osteoarthritis
XLH	X-chromosomale Hypophosphatämie

## Hintergrund

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

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

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

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

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Burosumab in seiner Sitzung am 26. Januar 2021 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 29. Oktober 2020 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pharmazeutischen Unternehmers in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 01. Februar 2021 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

## 1 Fragestellung

Burosumab (CRYSVITA®) ist ein Arzneimittel mit dem Status „Orphan Drug“. Gemäß § 35a Abs. 1 Satz 11 1. Halbs. SGB V gilt für Orphan Drugs der medizinische Zusatznutzen bereits durch die Zulassung als belegt. Der G-BA bestimmt bei Orphan Drugs, die einen Umsatz von 50 Millionen Euro in den letzten 12 Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien.

Die Zulassung von Burosumab umfasste bisher die Behandlung von Kindern ab 1 Jahr und Jugendlichen in der Skelettwachstumsphase mit X-chromosomaler Hypophosphatämie (XLH) und röntgenologischem Nachweis einer Knochenerkrankung [13].

Es erfolgte am 30. September 2020 die folgende Zulassungserweiterung:

- CRYSVITA wird angewendet zur Behandlung der XLH bei Kindern und Jugendlichen im Alter von 1 bis 17 Jahren mit röntgenologischem Nachweis einer Knochenerkrankung, und bei Erwachsenen [11].

Die vorliegende Bewertung bezieht sich gemäß Zulassungserweiterung auf erwachsene Patientinnen und Patienten mit XLH.

Gemäß Fachinformation [11] beträgt die empfohlene Anfangsdosis bei Erwachsenen 1,0 mg/kg Körpergewicht, auf die näheren 10 mg auf- bzw. abgerundet, bis zu einer Höchstdosis von 90 mg alle 4 Wochen.

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studienname (Nummer)	Studie mit Dossier vom pU eingereicht	Studie vom pU als relevant für Ableitung des Zusatznutzens erachtet	Studie relevant für Nutzenbewertung	Gründe, warum Studie nicht relevant für Nutzenbewertung
<i>Studien zum Wirkstoff</i>				
UX023-CL303 <sup>1)</sup>	Ja	Ja	Ja	-
KRN23-INT-001 <sup>1)</sup>	Nein	Nein	Nein	Es handelt sich um eine offene Dosissteigerungsstudie der Phasen I/II mit N = 32. Burosumab wurde in Dosen unterhalb der empfohlenen Anfangsdosis der FI verabreicht. Mit der Studie UX023-CL303 liegt Evidenz mit einer Kontrollgruppe vor.
KRN23-INT-002 <sup>1)</sup>	Nein	Nein	Nein	Es handelt sich um eine offene Extensionsstudie der Phasen I/II mit N = 23, in der auch Dosen unterhalb der empfohlenen Anfangsdosis der FI verabreicht wurden. Mit der Studie UX023-CL303 liegt Evidenz mit einer Kontrollgruppe vor.
UX023-CL203 <sup>1)</sup>	Nein	Nein	Nein	Es handelt sich um eine offene Extensionsstudie der Phase IIb mit N = 20, in der auch Dosen unterhalb der empfohlenen Anfangsdosis der FI verabreicht wurden. Mit der Studie UX023-CL303 liegt Evidenz mit einer Kontrollgruppe vor.
UX023-CL304 <sup>1)</sup>	Nein	Nein	Nein	Es handelt sich um eine offene Studie der Phase III mit N = 14. Mit der Studie UX023-CL303 liegt Evidenz mit einer Kontrollgruppe vor.
BUR-GER	Nein	Nein	Nein	Für diese einarmige Studie liegt kein Datenschnitt vor.

<sup>1)</sup> Zulassungsrelevante Studie gemäß EPAR [10].

Abkürzungen: EPAR: European Public Assessment Report; FI: Fachinformation; pU: pharmazeutischer Unternehmer

### Zur Nutzenbewertung für Burosumab wurden folgende Studien und Daten herangezogen:

- Herstellerdossier zu Burosumab [20],
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [10],
- Studienberichte inkl. Studienprotokoll und statistischem Analyseplan der Studie UX023-CL303 [21,34,35],
- Publikationen zum Wirkstoff: Insogna et al. 2018 [17] und Portale et al. 2019 [29].

## 2.2 Aufbau und Design der eingeschlossenen Studie

Die Unterlagen zur Nutzenbewertung für Burosumab basieren auf der Zulassungsstudie UX023-CL303 (im Folgenden als Studie CL303 bezeichnet). Die Studie und die Intervention werden in Tabelle 2 und Tabelle 3 charakterisiert.

Tabelle 2: Charakterisierung der Studie CL303

Charakteristikum	Beschreibung
<b>Design</b>	<p>Bei der Studie CL303 handelt es sich um eine randomisierte, doppelblinde, Placebo-kontrollierte, multizentrische Studie der Phase III zur Untersuchung der Wirksamkeit und Sicherheit von Burosumab bei Erwachsenen mit XLH. Primäres Zielkriterium der Studie war das Erreichen eines mittleren Serumphosphatspiegels oberhalb der unteren Grenze des Normalbereichs.</p> <p>Die Randomisierung erfolgte im Verhältnis 1:1 auf Burosumab (1,0 mg/kg Körpergewicht) oder Placebo und war stratifiziert nach den folgenden Faktoren:</p> <ul style="list-style-type: none"> <li>• Durchschnittliche Schmerzen in den letzten 7 Tagen (<i>Brief Pain Inventory-Short Form</i> [BPI-SF] Item 5 <math>\leq 6,0</math> vs. <math>&gt; 6,0</math>) <sup>1)</sup> und</li> <li>• Region (Nordamerika/Europa vs. Japan vs. Südkorea) <sup>2)</sup>.</li> </ul> <p>Die Studie besteht aus</p> <ul style="list-style-type: none"> <li>• einer Screeningperiode (einschließlich einer zweiwöchigen Phosphat- und Vitamin-D-Auswaschperiode),</li> <li>• einer verblindeten, Placebo-kontrollierten Behandlungsperiode (Woche 1 bis Woche 24),</li> <li>• einer offenen, einarmigen Behandlungsfortsetzungsphase (Woche 25 bis 48) und</li> <li>• einer freiwilligen, offenen und einarmigen Extensionsperiode I für 48 Wochen (Woche 49 bis Woche 96).</li> </ul> <p>In Südkorea und Japan stellt die Studienwoche 96 das Ende der Studie dar, wohingegen in Nordamerika/Europa die Behandlung mit Burosumab für bis zu 53 weitere Wochen bis September 2018 möglich war (Extensionsperiode II). Die maximale Studiendauer beläuft sich auf 149 Wochen.</p> <p>Die Studie ist abgeschlossen. Es liegt der finale Studienbericht vom 19. September 2019 mit Analysen bis Woche 24, 48, 96 und Studienende vor.</p>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Männer und Frauen im Alter von 18–65 Jahren</li> <li>• XLH-Diagnose, basierend auf typischen klinischen Merkmalen der XLH bei Erwachsenen (wie z. B. Kleinwuchs oder gebogene Beine) und mindestens eines der folgenden Merkmale beim Screening:</li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>○ Dokumentierte <i>PHEX</i>-Mutation entweder bei der Testperson oder bei einem direkt verwandten Familienmitglied mit entsprechender X-chromosomaler Vererbung</li> <li>○ iFGF23-Spiegel im Serum &gt; 30 pg/mL nach Kainos-Assay</li> <li>• Biochemische Befunde, die mit XLH bei Screeningvisite 2 nach nächtlichem Fasten (mind. 8 Stunden) übereinstimmen: <ul style="list-style-type: none"> <li>○ Serumphosphat &lt; 2,5 mg/dL (0,81 mmol/L)</li> <li>○ Tubuläres Maximum der Phosphatrückresorption (TmP/GFR) &lt; 2,5 mg/dL</li> </ul> </li> <li>• Vorhandensein von Knochenschmerzen, die auf eine XLH/Osteomalazie zurückzuführen sind, definiert durch einen Wert von <math>\geq 4</math> im Item 3 des BPI-SF („Schlimmste Schmerzen“) bei Screeningvisite 1</li> <li>• Geschätzte glomeruläre Filtrationsrate (eGFR) <math>\geq 60</math> mL/min (unter Verwendung der Formel der Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI]); oder eGFR von 45 bis &lt; 60 mL/min bei Screeningvisite 2 mit Nachweis, dass die Niereninsuffizienz nicht auf eine Nephrokalzinose zurückzuführen ist</li> <li>• Bei Einnahme von Schmerzmitteln für chronische Schmerzen (einschließlich Narkotika/Opioiden) müssen diese mindestens 21 Tage vor der ersten Screening-Visite in einem stabilen Dosierungsschema verabreicht werden und es muss die Bereitschaft bestehen, das Dosierungsschema während der gesamten Placebo-kontrollierten Behandlungsdauer beizubehalten. Die Dosis darf 60 mg orale Morphinäquivalente/Tag nicht überschreiten.</li> <li>• Ausfüllen des Patiententagebuchs an mindestens 4 von 7 Tagen vor der Baselinevisite</li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Verwendung eines pharmakologischen Vitamin-D-Metaboliten oder -Analogons (Calcitriol, Doxercalciferol und Paricalcitol) innerhalb von 14 Tagen vor Screeningvisite 2</li> <li>• Verwendung von oralem Phosphat innerhalb von 14 Tagen vor Screeningvisite 2</li> <li>• Verwendung von Aluminiumhydroxid-Antaziden, Acetazolamiden und Thiaziden innerhalb von 7 Tagen vor Screeningvisite 2</li> <li>• Chronischer Gebrauch von systemischen Kortikosteroiden, definiert als mehr als 10 Tage in den 2 Monaten vor Screeningvisite 1</li> <li>• Korrigierter Serumkalziumspiegel <math>\geq 10,8</math> mg/dL (2,7 mmol/L) bei Screeningvisite 2</li> <li>• intaktes Parathormon (iPTH) im Serum <math>\geq 2,5</math> ULN bei Screeningvisite 1</li> <li>• Einsatz von Arzneimitteln zur Suppression von Parathormon (PTH) (z. B. Cinacalcet) innerhalb von 60 Tagen vor Screeningvisite 1</li> <li>• Verwendung von Bisphosphonaten in den 2 Jahren vor Screeningvisite 1</li> <li>• Verwendung von Denosumab in den 6 Monaten vor Screeningvisite 1</li> <li>• Verwendung von Teriparatid in den 2 Monaten vor Screeningvisite 1</li> <li>• Geplanter oder empfohlener orthopädischer Eingriff innerhalb der ersten 24 Wochen der klinischen Studienphase</li> <li>• Auftreten einer traumatischen Fraktur oder eines orthopädischen Eingriffs innerhalb von 6 Monaten vor Screeningvisite 1</li> <li>• Verwendung des Prüfpräparats oder eines anderen therapeutischen monoklonalen Antikörpers innerhalb von 90 Tagen vor Screeningvisite 1</li> <li>• Verwendung eines Prüfpräparats innerhalb von 30 Tagen vor Screeningvisite 1 bzw. Bedarf eines Prüfpräparats vor Abschluss aller geplanten Studienuntersuchungen (in Japan: 4 Monate vor dem Screening)</li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>• Nicht möglich oder nicht bereit, untersagte Arzneimittel während der gesamten Studie zurückzustellen</li> <li>• Anamnese wiederkehrender Infektionen (mit Ausnahme von Zahnabszessen, von denen bekannt ist, dass sie mit XLH assoziiert sind) oder einer Prädisposition für Infektionen oder einer bekannten Immunschwäche</li> </ul>
<b>Intervention und Zahl der Patientinnen und Patienten</b>	<p>Es wurden 163 Testpersonen gescreent; davon wurden 134 randomisiert. Diese 134 Personen bilden die FAS-Population <sup>3)</sup>.</p> <p>Behandlungsgruppen:</p> <ul style="list-style-type: none"> <li>• Interventionsgruppe Burosumab: 68 Patientinnen und Patienten</li> <li>• Kontrollgruppe Placebo: 66 Patientinnen und Patienten</li> </ul>
<b>Ort und Zeitraum der Durchführung</b>	<p><b>Ort der Durchführung</b></p> <p>Die Studie wurde an 25 Studienzentren durchgeführt, darunter 8 in den Vereinigten Staaten (USA), je 5 im Vereinigten Königreich und in Japan, 3 in Frankreich, 2 in Südkorea und je 1 in Irland und Italien.</p> <p><b>Studiendauer</b></p> <ul style="list-style-type: none"> <li>• Erste Visite erste/erster Patientin/Patient: 22. Oktober 2015</li> <li>• Erste Visite letzte/letzter Patientin/Patient: keine Angabe</li> <li>• Letzte Visite letzte/letzter Patientin/Patient: 6. Dezember 2018</li> </ul>
<b>Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte gemäß Studienprotokoll</b>	<p><b>Primärer Endpunkt</b></p> <p>Erreichen von Serumphosphatwerten über der unteren Grenze des Normalwerts (LLN; 2,5 mg/dL [0,81 mmol/L]) in der Mitte des Dosiszyklus (d. h. in den Wochen 2, 6, 10, 14, 18 und 22)</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Veränderung in BPI-SF Item 3 („Schlimmste Schmerzen“)</li> <li>• Veränderung im <i>Western Ontario and McMaster Universities Osteoarthritis Index</i> (WOMAC)-Wert „Steifigkeit“</li> <li>• Veränderung im WOMAC-Wert „Körperliche Funktion“</li> <li>• Abnahme um mind. 15 % und mind. 30 % in BPI-SF Frage 3</li> <li>• Abnahme um mind. 10 <i>Normalised Units</i> (nu) im WOMAC-Wert „Steifigkeit“</li> <li>• Abnahme um mind. 9,3 nu im WOMAC-Wert „Körperliche Funktion“</li> <li>• Erreichen von Serumphosphatwerten über LLN zum Ende eines Dosiszyklus (d. h. 4 Wochen nach der letzten Dosis)</li> <li>• Veränderung im Serumphosphat in der Mitte und zum Ende der Dosiszyklen</li> <li>• Veränderung im 1,25-Dihydroxyvitamin D im Serum, Phosphat im Urin, Renale Phosphatschwelle (TmP/GFR), prozentuale tubuläre Phosphatrückresorption (TRP) und in biochemischen Markern des Knochenumbaus, einschließlich Prokollagen Typ-1-N-Propeptid (P1NP), C-terminales Kollagen Typ I-Telopeptid-Quer (cross)-Vernetzung (CTx) und knochenspezifische alkalische Phosphatase (BALP)</li> <li>• Veränderungen in BPI-SF Schmerzintensität und Schmerzbeeinträchtigung</li> <li>• Veränderung in <i>Brief Fatigue Inventory</i> (BFI) Frage 3 („Schlimmste Fatigue“)</li> <li>• Veränderung im Gesamtwert des BFI</li> </ul> <p><b>Explorative Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Anzahl der aktiven Pseudofrakturen und/oder Frakturen</li> <li>• Veränderung der Belastung durch Fersenbeinenthesopathie</li> <li>• Veränderung im <i>Patient Global Impression of Improvement</i> (PGI-I)</li> <li>• Veränderung im 6-Minuten-Gehtest (6MWT)</li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>Veränderung im <i>Timed Up and Go</i> (TUG)-Test</li> </ul> <p><b>Sicherheit</b> Inzidenz der UE und SUE</p>
<b>Subgruppenanalysen</b>	<p>Im SAP waren für die Wirksamkeitsendpunkte die folgenden Subgruppenanalysen a priori geplant:</p> <ul style="list-style-type: none"> <li>BPI-SF „Schlimmster Schmerz“ zu Baseline (<math>\leq 6,0</math> vs. <math>&gt; 6,0</math>)</li> <li>BPI-SF „Durchschnittlicher Schmerz“ zu Baseline (<math>\leq 6,0</math> vs. <math>&gt; 6,0</math>)</li> <li>Region (Nordamerika/EU vs. Japan vs. Südkorea)</li> <li>Geschlecht (weiblich vs. männlich)</li> <li>Ethnie (weiß vs. nicht weiß).</li> </ul> <p>Laut SAP sollten für die Sicherheitsanalysen ebenfalls getrennte Darstellungen nach Region angefertigt werden.</p> <p>Zusätzlich im Studienbericht sind folgende Subgruppenanalysen dargestellt:</p> <ul style="list-style-type: none"> <li>Pseudofrakturen zu Baseline (ja vs. nein)</li> <li>WOMAC „Steifheit“ zu Baseline (<math>\leq</math> Median vs. <math>&gt;</math> Median)</li> <li>WOMAC „Physische Funktion“ zu Baseline (<math>\leq</math> Median vs. <math>&gt;</math> Median).</li> </ul> <p>Der pU präsentiert im Dossier Modul 4 zusätzlich diese Subgruppenanalysen:</p> <ul style="list-style-type: none"> <li>Alter (<math>\leq 41</math> Jahre vs. <math>&gt; 41</math> Jahre)</li> <li>Verwendung von Schmerzmitteln zu Baseline (ja vs. nein)</li> <li>Verwendung von Opioiden zu Baseline (ja vs. nein)</li> <li>WOMAC „Schmerzen“ zu Baseline (<math>\leq 50</math> vs. <math>&gt; 50</math>)</li> <li>WOMAC Gesamtwert zu Baseline (<math>\leq 49,5</math> vs. <math>&gt; 49,5</math>)</li> <li>6MWT zu Baseline (<math>\leq 365</math> m vs. <math>&gt; 365</math> m).</li> </ul>

<sup>1)</sup> Ursprünglich war eine Stratifizierung für „Schlimmste Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 3) vorgesehen. Im IWRS wurde fälschlicherweise eine Stratifizierung für Item 5 des BPI-SF programmiert. Da die Rekrutierung bereits begonnen hatte, als der Fehler bemerkt wurde, wurde die stratifizierte Randomisierung mit Item 5 fortgeführt.

<sup>2)</sup> Der Faktor „Region“ wurde mit Amendment 3 des Studienprotokolls eingeführt.

<sup>3)</sup> Als FAS-Population waren alle Patientinnen und Patienten definiert, die mind. eine Dosis der Studienmedikation während der Placebo-kontrollierten Behandlungsperiode erhielten. Da dies auf alle randomisierten Patientinnen und Patienten der Studie UX023-CL303 zutrifft, entspricht die FAS-Population der ITT-Population.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; BFI: Brief Fatigue Inventory; BPI-SF: Brief Pain Inventory-Short Form; dL: Deziliter; FAS: Full Analysis Set; ITT: Intention to treat; LLN: untere Grenze des Normalwerts; mg: Milligramm; mmol: Millimol; L: Liter; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; (S)UE: (Schwerwiegendes) unerwünschte(s) Ereignis(se); vs: versus; WOMAC: Western Ontario and McMaster Universities Osteoarthritis; XLH: X-chromosomale Hypophosphatämie

**Tabelle 3: Charakterisierung der Interventionen in der Studie CL303 während der Placebo-kontrollierten Behandlungsperiode**

Intervention	Kontrolle
<p><b>Burosumab</b></p> <ul style="list-style-type: none"> <li>Dosierung: <ul style="list-style-type: none"> <li>1,0 mg/kg Körpergewicht alle 4 Wochen (max. Gesamtdosis: 90 mg)</li> <li>Die Dosis sollte während der gesamten Studiendauer konstant bleiben unter der Voraussetzung, dass der Phosphat Spiegel im Serum zu keinem Zeitpunkt 5,0 mg/dL (1,61 mmol/L) oder zweimal 4,5 mg/dL (1,45 mmol/L) überschritt und sich das Körpergewicht um nicht mehr als 20 % gegenüber der Ausgangsmessung</li> </ul> </li> </ul>	<p><b>Placebo</b></p> <ul style="list-style-type: none"> <li>Zusammensetzung: <ul style="list-style-type: none"> <li>Sterile, klare, farblose und konservierungsmittelfreie Lösung in 5-mL-Einwegfläschchen mit 1 mL Placebo</li> <li>Gleiche Zusammensetzung wie beim Prüfpräparat Burosumab, außer dass Placebo keinen Wirkstoff enthielt</li> </ul> </li> <li>Applikation: In gleicher Weise wie Burosumab vorbereitet und verabreicht, wobei die gleichen Richtlinien für gewichtsbezogene</li> </ul>

Intervention	Kontrolle
<p>veränderte. Die Dosis wurde neu berechnet, wenn sich das Körpergewicht um mehr als 20 % änderte.</p> <ul style="list-style-type: none"> <li>○ Bei Überschreiten von 5,0 mg/dL (1,61 mmol/L) Serumphosphat sollte die Verblindung aufgehoben und die Dosis um die Hälfte reduziert werden.</li> <li>○ Bei Überschreiten der oberen Grenze des Normalbereichs (ULN) von 4,5 mg/dL (1,45 mmol/L) Serumphosphat sollte die Verblindung nur aufgehoben werden, wenn auch die zweite Messung oberhalb der ULN lag.</li> </ul> <ul style="list-style-type: none"> <li>● Applikation: Subkutane Injektion in Bauch, Oberarme, Oberschenkel oder Gesäß (Injektionsstelle wechselt bei jeder Injektion). Übersteigt die Dosis ein Volumen von 1,5 ml, sollte die Dosis an zwei Injektionsstellen verabreicht werden.</li> </ul>	<p>Berechnungen, Annahmen zur maximalen Dosis, Injektionsmethode und zur Entblindung verwendet wurden.</p>
<p><b>Nicht erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>● Pharmakologische Vitamin-D-Metaboliten oder -Analoge (z. B. Calcitriol, Doxercalciferol und Paricalcitol)</li> <li>● Orales Phosphat</li> <li>● Aluminiumhydroxid-Antazide, Acetazolamid oder Thiazid</li> <li>● Bisphosphonate</li> <li>● Denosumab</li> <li>● Teriparatid</li> <li>● Chronischer Gebrauch von systemischen Kortikosteroiden (kurze Verabreichungszeiten sind akzeptabel, falls angezeigt)</li> <li>● Parathormon-Suppressoren (z. B. Cinacalcet; 60-Tage-Auswaschung erforderlich)</li> <li>● Jede andere Therapie mit einem monoklonalen Antikörper (außer Studienmedikation)</li> </ul> <p><b>Erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>● Orale Supplementation von 25-Hydroxyvitamin D möglich bei unter 20 ng/mL im Serum</li> <li>● Nichtsteroidale Antirheumatika, Opioide oder andere narkotische Schmerzmittel, die in einer stabilen Dosis verwendet wurden, waren zur Schmerzkontrolle zugelassen (begrenzt auf 60 mg orales Morphin-Äquivalent pro Tag). Eine Änderung der Anwendung oder Dosis von narkotischen Schmerzmitteln/Opioiden zur Kontrolle von Schmerzen im Zusammenhang mit einem akuten medizinischen Ereignis war möglich.</li> </ul>	

Abkürzungen: dL: Deziliter; kg: Kilogramm; L: Liter; mL: Milliliter; mg: Milligramm; mmol: Millimol; ng: Nanogramm; ULN: obere Grenze des Normalbereichs

### Vorgelegte Studienberichte und Analysezeiträume

Auswertungen der Studie CL303 liegen für die folgenden Auswertungszeiträume vor:

- die Placebo-kontrollierte Behandlungsperiode (Woche 1 bis Woche 12; Woche 1 bis Woche 24),
- zusätzlich mit der offenen, einarmigen Behandlungsfortsetzungsphase (Woche 1 bis 48),
- zusätzlich mit der offenen, einarmigen Extensionsperiode I (Woche 1 bis Woche 96),
- zusätzlich mit der offenen, einarmigen Extensionsperiode II (Woche 1 bis Woche 120; Woche 1 bis Studienende).

Es wurden 3 konsekutive Studienberichte vorgelegt. Der erste Studienbericht Woche 24 [34] berichtet Ergebnisse der Placebo-kontrollierten Behandlungsperiode, der zweite Studienbericht Woche 48 [35] enthält zusätzlich Auswertungen zur offenen, einarmigen Behandlungsfortsetzungsphase, der Studienbericht Woche 96 [21] umfasst zusätzlich Ergebnisse aus den beiden einarmigen Extensionsperioden. Gruppenunterschiede wurden ausschließlich für die Placebo-kontrollierte Behandlungsperiode berechnet. Für die Nutzenbewertung werden aufgrund der vorhandenen Kontrollgruppe Ergebnisse dieser Behandlungsperiode, d. h. für Woche 1 bis Woche 24, herangezogen. Für das Dossier reichte der pU zudem ein Nachberechnungsdokument ein, welches erneute Auswertungen zu allen Endpunkten des Modul 4 enthält.

### Protokolländerungen

Das Originalstudienprotokoll ist auf den 5. Juni 2015 datiert. Es wurde durch die 5 Protokollamendments in Tabelle 3 modifiziert. Relevante Änderungen sind ebenfalls in Tabelle 3 dargestellt.

Der Einschluss des ersten Patienten / der ersten Patientin ist auf den 22. Oktober 2015 datiert. Der Datenschnitt für die erste Auswertung zum Abschluss der Placebo-kontrollierten Behandlungsperiode wurde am 22. Dezember 2016 durchgeführt, sodass die Rekrutierung vor diesem Datum abgeschlossen war.

Tabelle 4: Übersicht der Protokollamendments

Amendment	Anzahl der bis dahin eingeschlossenen Patientinnen und Patienten	Relevante Änderungen
<b>Amendment 1 vom 21. Juli 2016</b>	Keine Angabe <sup>1)</sup>	<p>Das Studiendesign wurde um eine 48-wöchige Extensionsperiode erweitert.</p> <p>Als Ausschlusskriterien wurden die Einnahme von Denosumab in den 6 Monaten und die Einnahme von Teriparatid in den 2 Monaten vor der Screeningvisite ergänzt. Die Arzneimittel wurden als nicht erlaubte Begleitmedikation ergänzt.</p> <p>Die Erhebung des TUG-Tests ab Woche 24 wurde in der Studie ergänzt. Der Endpunkt wurde als explorativ definiert.</p> <p>Die Erhebung der Fersenbeinenthesopathie ab Woche 24 wurde als explorativer Wirksamkeitsendpunkt ergänzt.</p>
<b>Amendment 2 vom 8. September 2016</b>	Keine Angabe <sup>1)</sup>	Keine
<b>Amendment 3 vom 31. März 2017</b>	Alle Personen wurden vor diesem Amendment eingeschlossen.	<p>Als Stratifizierungsfaktor wurde Region (Nordamerika/Europa, Japan und Südkorea) zusätzlich aufgenommen.</p> <p>Eine zweite Extensionsperiode für ca. 53 weitere Wochen wurde für Studienteilnehmende in den USA hinzugefügt.</p>
<b>Amendment 4 vom 29. September 2017</b>	Alle Personen wurden vor diesem	Die zweite Extensionsperiode wurde auch für Studienteilnehmende aus Europa geöffnet.

Amendment	Anzahl der bis dahin eingeschlossenen Patientinnen und Patienten	Relevante Änderungen
	Amendment eingeschlossen.	Die Erhebungen des BPI-SF und BFI wurden nach Woche 96 nicht fortgesetzt.
<b>Amendment 5 ohne Datumsangabe</b>	Alle Personen wurden vor diesem Amendment eingeschlossen.	Keine

<sup>1)</sup> Mit erstem Datenschnitt zu Woche 24 am 22. Dezember 2016 war die Rekrutierung vor diesem Datum abgeschlossen. Das tatsächliche Datum des Einschlusses der letzten Testperson ist jedoch nicht angegeben.

Abkürzungen: BFI: Brief Fatigue Inventory; BPI-SF: Brief Pain Inventory-Short Form; TUG: Timed Up and Go-Test

Hinsichtlich der Änderungen zu Denosumab und Teriparatid aus Amendment 1 ist unklar, wie mit bis dahin bereits eingeschlossenen Studienteilnehmenden umgegangen wurde, die eine solche Therapie erhielten.

Die Umsetzung der Änderung der stratifizierten Randomisierung mit Amendment 3 ist ebenfalls nicht nachvollziehbar. Im ersten Studienbericht mit Datenschnitt vor Amendment 3 waren bereits alle Studienteilnehmenden rekrutiert.

## 2.3 Endpunkte

In diesem Kapitel werden die Eignung der Endpunkte hinsichtlich Operationalisierung, Patientenrelevanz und Validität beurteilt. Dazu wurden das Herstellerdossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt. Tabelle 5 stellt das Ergebnis dieser Bewertung zusammenfassend dar.

Tabelle 5: Zusammenfassung der Endpunktbewertung

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Gesamtmortalität <sup>1)</sup>	Mortalität	Ja	Ja
Serumphosphat <sup>2)</sup>	Morbidität	Ja	Ergänzend
Gehfähigkeit (6MWT)		Ja	Ja
Schmerzen, Steifheit und „Physische Funktion“ (WOMAC)		Ja	Ja
Schmerzen (BPI-SF)		Ja	Ja
Fatigue (BFI)		Ja	Ja
Patientenberichtete Verbesserung (PGI-I)		Ja	Ja
Unerwünschte Ereignisse	Sicherheit	Ja	Ja

<sup>1)</sup> Im Rahmen der Sicherheit erfasst.

<sup>2)</sup> Primärer Endpunkt.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; BFI: Brief Fatigue Inventory; BPI-SF: Brief Pain Inventory - Short Form; PGI-I Patient Global Impression of Improvement; pU: pharmazeutischer Unternehmer; WOMAC: Western Ontario and McMaster Universities Osteoarthritis

## 2.3.1 Mortalität

### Gesamtmortalität

Der Endpunkt Gesamtmortalität wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung:

##### *Beschreibung*

Die Gesamtmortalität umfasst die Anzahl der Patientinnen und Patienten, die im Studienverlauf im Rahmen der Sicherheitserfassung als Todesfälle gemeldet wurden.

Die Erhebung wurde gemäß Medical Dictionary for Regulatory Activities (MedDRA), v18.1. durchgeführt. Es wurden alle UE ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung bis zur letzten Studienvisite / Visite bei vorzeitigem Abbruch aufgezeichnet. Eine anschließende telefonische Nachbeobachtung nach 4 und 8 Wochen fand nur bei Patientinnen und Patienten statt, die nicht unmittelbar ihre Burosumab-Therapie durch Marktzulassung oder auf anderem Wege fortsetzten.

Im SAP findet sich die Einschränkung, dass nur UE unter Behandlung (Treatment Emergent Adverse Events, TEAE) ausgewertet werden, d. h. UE, die bei oder nach der ersten Dosis des Prüfprodukts auftraten und nicht bereits vor der ersten Dosisgabe bestanden.

- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu jeder Studienvisite erhoben. Bei Patientinnen und Patienten, die nicht unmittelbar ihre Burosumab-Therapie durch Marktzulassung oder auf anderem Wege nach Studienende fortsetzten, fand eine anschließende telefonische Nachbeobachtung nach 4 und 8 Wochen statt.
- Auswertung:  
Die Auswertung erfolgte auf Basis der Safety-Population mit Datenschnitt zu Woche 24.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz:

Bei Gesamtmortalität handelt es sich um einen patientenrelevanten Endpunkt.

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

#### Validität:

In der Studie CL303 werden UE ab Unterzeichnung der Einverständniserklärung erfasst, jedoch erst ab der ersten Dosis der jeweiligen Studienmedikation ausgewertet. Es ist unklar, wieso UE erhoben, jedoch nicht ausgewertet werden.

## 2.3.2 Morbidität

### Serumphosphat

Der Endpunkt Serumphosphat wird in der Nutzenbewertung ergänzend dargestellt.

#### Operationalisierung:

##### *Beschreibung*

Die Bestimmung von Serumphosphat war durch ein Studienmanual für alle Studienzentren standardisiert.

Zur Entnahme von Serumproben mussten die Studienteilnehmenden seit mindestens 8 Stunden (über Nacht) nüchtern sein. Die Entnahme sollte nach Möglichkeit morgens und bei jeder Visite etwa zur gleichen Zeit stattfinden. Die Blutproben wurden vor der Verabreichung des Studienmedikaments entnommen und für alle Studienvisiten ab Baseline zentral und verblindet ausgewertet.

Der pU bietet Auswertungen der mittleren Differenzen in den Änderungen von Baseline zu Woche 24 zwischen den Behandlungsgruppen mit einem post hoc berechneten Hedges g an. Zusätzlich wertet der pU die Anteile in den Behandlungsgruppen aus, die einen Serumphosphat-Wert oberhalb der unteren Grenze des Normalbereichs in der Mitte des Dosiszyklus und zum Ende eines Dosiszyklus haben.

- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu Baseline, Woche 2, Woche 4, Woche 6, Woche 10, Woche 12, Woche 14, Woche 18, Woche 20, Woche 21, Woche 22, Woche 24 und vorzeitiger Abbruch / Behandlungsende erhoben.
- Auswertung:  
Die Auswertung erfolgte auf Basis der FAS-Population mit Datenschnitt zu Woche 24.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Der pU diskutiert in seinem Dossier jedoch nicht, inwiefern eine Responder-Definition als Werte im Normalbereich gegenüber der gewählten Definition oberhalb der unteren Grenze des Normalbereichs vorzuziehen gewesen wäre.

#### Patientenrelevanz:

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

Als Laborparameter stellt Serumphosphat keinen unmittelbar patientenrelevanten Endpunkt dar. Der pU hat keine Nachweise vorgelegt, die die Validität als Surrogat für patientenrelevante Endpunkte der Morbidität oder Lebensqualität begründen.

Die Erhöhung des pathologisch erniedrigten Serumphosphatwerts bis zum Erreichen des Normalbereichs ist als Therapieziel jedoch der klinisch bedeutsame Parameter. Der pU legt eine entsprechende Auswertung vor, welche daher ergänzend dargestellt wird. Dabei wird die Response zum Ende des Dosiszyklus als geeigneter eingeschätzt, eine nachhaltige Response abzubilden. Da diese Auswertung jedoch ausschließlich deskriptiv erfolgte, wird ebenfalls die

Response in der Mitte des Dosiszyklus abgebildet. Der pU diskutiert die Relevanz der beiden Intervalle in seinem Dossier nicht.

#### Validität:

Die Bestimmung von Serumphosphat fand in einem zentralen Labor verblindet statt und wird daher als valide bewertet.

#### **6-Minuten-Gehtest (6MWT)**

Der Endpunkt Gehfähigkeit mittels 6MWT wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung:

##### *Beschreibung*

Zur Messung der körperlichen Funktion wurde in der Studie CL303 der 6MWT eingesetzt.

Laut Studienunterlagen wurde der Test nach den Vorgaben der *American Thoracic Society* (ATS) [1] von geschultem Studienpersonal durchgeführt. Testpersonen wurden angewiesen, eine vorher abgemessene Strecke über 6 Minuten (wiederholt) zu laufen. Die zurückgelegte Strecke wurde nach Ablauf der 6 Minuten in Metern erfasst. Funktionstests sollten nach Ausfüllen der PRO-Instrumente in den Studienvisiten stattfinden.

Bei Personen, die zur ersten Screeningvisite nicht in der Lage waren, den Test durchzuführen, wurde auf die Testdurchführung im weiteren Studienverlauf verzichtet.

Laut elektronischem Prüfbogen (eCRF) waren Orthesen und andere Gehhilfen in der Durchführung erlaubt. Weitere Informationen zur standardisierten Durchführung in den Studienzentren (z. B. aus einem Manual) wurden nicht identifiziert.

Der pU stellt neben Auswertungen der zurückgelegten Gehstrecke als Gruppenunterschied der mittleren Veränderungen von Baseline zu Woche 24 zwischen den Studienarmen zusätzlich eine Auswertung als Gruppenunterschied der mittleren Veränderungen von Baseline zu Woche 24 zwischen den Studienarmen im Vergleich zu Normwerten dar, die aus einer Untersuchung von Gibbons et al. [14] stammen.

- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu Baseline, Woche 12 und Woche 24 und bei vorzeitigem Abbruch erhoben.
- Auswertung:  
Die Auswertung erfolgte auf Basis der FAS-Population mit Datenschnitt zu Woche 24.

##### *Bewertung*

Die Operationalisierung des Endpunkts entspricht gemäß pU den Vorgaben der ATS. Die konkrete Umsetzung dieser Richtlinienvorgaben (z. B. hinsichtlich Testinstruktion, Schulung der Testdurchführenden, Probelauf) im Rahmen der Studie CL303 bleibt unklar.

#### Patientenrelevanz:

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

Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Die Gehfähigkeit erhoben mittels 6MWT als eine zentrale körperliche Funktion des alltäglichen Lebens ist geeignet zur Erfassung der körperlichen Belastbarkeit und stellt einen patientenrelevanten Endpunkt dar.

#### Validität:

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 XLH-Erkrankten erzielt werden können. Die Reproduzierbarkeit der Testergebnisse des 6MWT ist bei qualitätssichernden Maßnahmen hoch. Zu diesen Maßnahmen gehören z. B. Schulungen des Studienpersonals, standardisierte Formulierungen der Ermutigungen nach bestimmten Zeiteinheiten, Messung durch die gleiche Person, ein bis zwei Probetestläufe.

In der ATS-Richtlinie werden mehrere potentielle Einflussfaktoren bzw. Quellen für Variation der Testergebnisse des 6MWT genannt. Hierzu zählen zum einen Patientencharakteristika (Alter, Körpergröße und -gewicht, Geschlecht) als auch äußere Faktoren wie Motivation, Schulung des Studienpersonals und Erfahrung der Testpersonen mit dem 6MWT. Wird eine Veränderung der körperlichen Fähigkeiten durch den 6MWT ermittelt, sollten mögliche Einflussfaktoren konstant gehalten werden, um eine potentielle Veränderung der Gehstrecke auf den zu untersuchenden Faktor, in diesem Fall auf die Studienmedikation, zurückführen zu können.

Der vom pU vorgeschlagene Normwertvergleich (Prozentsatz der vorhergesagten Normalstrecke) wird als nicht valide gesehen, da die Repräsentativität der zugrunde liegenden Normstichproben aufgrund der willkürlichen Samplingstrategie (*Convenience Sample*) [14] nicht gegeben ist.

#### **Western Ontario and McMaster Universities Osteoarthritis (WOMAC)-Fragebogen**

Die Endpunkte „Steifheit“, „Physische Funktion“ und „Schmerzen“ mittels WOMAC werden in der Nutzenbewertung berücksichtigt.

#### Operationalisierung:

##### *Beschreibung*

Der WOMAC ist ein Instrument zur Beurteilung von Schmerzen, Steifheit und physischer Funktionsfähigkeit bei Patientinnen und Patienten mit Hüft- und/oder Kniearthrose.

Der WOMAC besteht aus 24 Items, die in 3 Subskalen unterteilt sind:

- Schmerzen (5 Items),
- Steifheit (2 Items),
- Physische Funktion (17 Items).

Der WOMAC ist verfügbar mit 3 Antwortformaten:

- Version LK3.1 mit 5-Punkte-Likert-Skalen,
- Version VA3.1 mit 100 mm visuellen Analogskalen und
- Version NR3.1 mit numerischen Analogskalen mit 11 Punkten.

In der Studie CL303 wurde die Version LK3.1 mit 5-Punkte-Likert-Skalen (von 0  $\triangleq$  keine bis 4  $\triangleq$  extrem) eingesetzt. Zur Auswertung stehen verschiedene Vorgehensweisen laut Manual [2] zur Verfügung, wobei vom pU die Bildung normalisierter Werte gewählt wurde: Dazu werden die

Itemwerte in jeder Subskala addiert, wobei folgende Werte in jeder Subskala erreichbar sind: Schmerz = 0 bis 20, Steifheit = 0 bis 8, Physische Funktion = 0 bis 68. Die Werte werden anschließend auf einer metrischen Skala von 0–100 (*normalised units*, nu) als Prozentwert des Höchstwerts skaliert. Das gleiche Vorgehen wird bei der Bildung des Gesamtwerts angewendet. Höhere Werte korrespondieren mit höheren Schmerzen, stärkerer Steifheit und eingeschränkterer körperlicher Funktion.

Es existieren zudem WOMAC-Versionen mit verschiedenen Bezugszeiträumen (von 24 Stunden bis letzter Monat). In der Studie CL303 wurde die Standardversion mit einem Bezugszeitraum von 48 Stunden verwendet.

PRO-Instrumente wurden in der folgenden Reihenfolge vor Durchführung der Funktionstests beantwortet: BPI-SF, BFI, WOMAC und PGI.

Es wurde in der Studie CL303 der vollständige WOMAC-Fragebogen mit 24 Items eingesetzt. Als Wirksamkeitsendpunkte waren lediglich die Subskalen „Steifheit“ und „Physische Funktion“ vorgesehen. Dies wird damit begründet, dass Schmerzen über den BPI-SF erfasst werden [33]. Die Subskalen wurden als Gruppenunterschied in den mittleren Veränderungen von Baseline zu Woche 24 und als Responderanalysen mit den Schwellenwerten 10 nu für die Subskala „Steifheit“ und 9,3 nu für die Subskala „Physische Funktion“ ausgewertet.

Laut COA-Dossier und Modul 4 wurden folgende Regeln für den Umgang mit fehlenden Werten angesetzt: Sofern beide Items der Subskala „Steifheit“, mind. 2 Items der Subskala „Schmerzen“ und mind. 4 Items der Subskala „Physische Funktion“ fehlten, wurde die Subskala als fehlend bewertet und nicht ausgewertet. Andernfalls wurde der Mittelwert aus den vorhandenen Items für fehlende Werte imputiert.

Der pU liefert mit seinem Dossier zudem Post-hoc-Auswertungen der Subskala „Schmerzen“ als Gruppenunterschied in den mittleren Veränderungen von Baseline zu Woche 24 zwischen den Studienarmen. Weiterhin wurde der Gruppenunterschied in den mittleren Veränderungen zwischen den Studienarmen im WOMAC-Gesamtwert in einer Post-hoc-Auswertung analysiert. Zur Einschätzung der klinischen Relevanz liefert der pU zudem Hedges g.

- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu Baseline, Woche 12 und Woche 24 erhoben.
- Auswertung:  
Die Auswertung erfolgte auf Basis der FAS-Population mit Datenschnitt zu Woche 24.

### *Bewertung*

Die Operationalisierung ist mit Einschränkungen nachvollziehbar. Die Rationale dafür, dass die Subskala Schmerzen erhoben wurde, aber nicht zur Auswertung vorgesehen war, kann nicht nachvollzogen werden.

Zudem ist keine Rationale verfügbar, warum das Messinstrument nicht in der Studienvsiste bei vorzeitigem Abbruch eingesetzt wurde.

### Patientenrelevanz:

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der

Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“.

Bei den Endpunkten Schmerzen, Steifheit und Physische Funktion handelt es sich um patientenrelevante Endpunkte.

#### Validität:

Der Fragebogen wurde unter Einbindung von Patientinnen und Patienten mit Hüft- und/oder Kniearthrose entwickelt. WOMAC wurde für diese Population ausgiebig getestet; die Reliabilität und Validität wurden nachgewiesen. Hinsichtlich der Änderungssensitivität zeigte sich jedoch die Subskala „Physische Funktion“ als weniger aussagekräftig als die anderen Subskalen, wobei die Eigenschaft grundsätzlich vorhanden zu sein scheint [3,7,23,32].

Zur Validität der WOMAC-Subskalen „Steifheit“ und „Physische Funktion“ hat der pU zudem eigene Untersuchungen für die Population der XLH-Erkrankten durchgeführt. Die Ergebnisse werden in einem *Clinical Outcome Assessment* (COA)-Dossier in einer Version 12 vom 18. September 2018 präsentiert [33]. Das COA-Dossier begründet übergreifend für das Entwicklungsprogramm zu Burosumab die Endpunktauswahl und -gestaltung, indem die Ergebnisse der durchgeführten Untersuchungen zusammenfassend dargestellt sind.

Für die Auswahl des WOMAC wird auf eine Interviewstudie mit 18 erwachsenen XLH-Erkrankten verwiesen, in der zunächst die Relevanz der Symptome Schmerzen, Steifheit und physische Funktion bestätigt wurde. Auch die Verständlichkeit der WOMAC-Subskalen „Steifheit“ und „Physische Funktion“ wurde überprüft. Obwohl die Fragen des WOMAC auf Osteoarthritis ausgerichtet sind, schränkte dies die Anwendbarkeit für die Befragten nicht ein.

Weitere psychometrische Untersuchungen wurden an Vorstudien der Studie CL303, jedoch auch anhand der Studie CL303 selbst durchgeführt. In der ersten Stufe der Untersuchungen wurden die Daten aus einer Querschnittserhebung (Studie UX023-CL001 mit N = 201), einer Phase-II-Studie (UX023-CL203 mit N = 20) und einer Phase-I-Studie (KRN23-INT-001) mit anschließender Extensionsstudie (KRN23-INT-002 mit N = 17 mit PRO-Daten) verwendet. Im Folgenden werden lediglich Ergebnisse dargestellt, die aus dieser ersten Stufe der psychometrischen Untersuchung und damit nicht aus der Studie CL303 stammen.

Hinsichtlich der Verteilung der Itemwerte zeigten sich für 6 der 17 Items der Subskala „Physische Funktion“ Bodeneffekte, da hohe Anteile bei den Antworten für geringe bis keine Einschränkungen festgestellt wurden. Für ein Item dieser Subskala wurde zudem ein Deckeneffekt beobachtet. Insgesamt konnten jedoch für die Mehrheit der Items gute Ergebnisse erzielt werden. Eine konfirmatorische Faktorenanalyse zeigte gute Ergebnisse hinsichtlich der Zuordnung der Items in den Subskalen, jedoch wiesen Parameter der Modellanpassung auf redundante Items hin. Die Known-Groups-Validität wurde nachgewiesen, indem anhand der Subskalen zwischen klinisch unterschiedlichen Gruppen unterschieden werden konnte. Konstruktvalidität wurde zudem über Korrelationen mit der Skala für „Physische Funktion“ der *Short Form* (SF)-36 mit positiven Ergebnissen untersucht. Die Stärke der Spearman-Korrelationen betrug -0,63 für die Subskala „Schmerzen“, -0,53 für die Subskala „Steifheit“ und -0,80 für die Subskala „Physische Funktion“. Die interne Konsistenz bewegte sich mit Cronbachs  $\alpha \geq 0,7$  mindestens im ausreichenden Bereich. Eine Untersuchung der Änderungssensitivität war aufgrund der geringen Anzahl an Personen mit Veränderungen nicht möglich [33].

Der pU untersuchte zudem klinische Relevanzschwellen für XLH-Erkrankte, indem er verteilungs- und ankerbasierte Methoden einsetzte. Die Ableitung der ankerbasierten MID fand auf Basis der Studie UX023-CL203 statt. Als Anker wurde eine Verbesserung um > 3 Punkte im Wert für „Physische Funktion“ im SF-36 zu Woche 24 verwendet. Dies entsprach 13 Personen in der Analyse zur Subskala „Schmerzen“ und jeweils 15 Personen in den Analysen der Subskalen „Steifheit“ und „Physische Funktion“. Die Ableitung der verteilungsbasierten MID wurde als 1/2 Standardabweichung (SD) zu Baseline und *Standard Error of Measurement* (SEM) durchgeführt. Es wurden folgende Ergebnisse erzielt:

- Subskala „Schmerzen“:
  - Ankerbasierte MID (SD): -8,08 (16,3)
  - Verteilungsbasierte MID: 7,24–11,28
- Subskala „Steifheit“:
  - Ankerbasierte MID (SD): -16,67 (21,48)
  - Verteilungsbasierte MID: 7,16–12,45
- Subskala „Physische Funktion“:
  - Ankerbasierte MID (SD): -13,63 (21,82)
  - Verteilungsbasierte MID: 4,10–12,35

Als Responderkriterium für die Auswertung in der Studie CL303 verwendet der pU die vorgeschlagenen Schwellenwerte aus dem WOMAC-Handbuch [2] von 10 für die Subskala „Steifheit“ und 9,3 für die Subskala „Physische Funktion“. Die Werte wurden in zwei identisch aufgebauten RCT mit insgesamt 1.545 Patientinnen und Patienten mit Hüft- und/oder Kniearthrose und den Interventionsarmen Rofecoxib, Ibuprofen und Placebo abgeleitet [9]. Als Anker wurde das patientenberichtete Ansprechen auf die Therapie nach 6 Wochen auf einer 5-Punkte-Skala von exzellentes Ansprechen bis kein Ansprechen verwendet. Der Bezugszeitraum von WOMAC ist nicht benannt. WOMAC kam in der Version VA3.1 mit 100 mm VAS zum Einsatz. Klinische Relevanzschwellen wurden definiert als die Differenz der mittleren Veränderung gegenüber dem Ausgangswert in WOMAC zwischen Befragten mit  $0 \triangleq$  „Kein Ansprechen“ und Befragten mit  $1 \triangleq$  „Schlechtes, unzureichendes Ansprechen“. Die Anzahl der Personen, die sich entsprechend einstuften, ist nicht angegeben. Für die Subskala „Steifheit“ entsprach diese einem Wert von 9,3 mm, für die Subskala „Physische Funktion“ einem Wert von 10,0 mm.

Laut Handbuch [2] wurde die Vergleichbarkeit der unterschiedlichen Antwortformate des WOMAC untersucht. Die Ergebnisse legen nahe, dass Werte der Version VA3.1 eine akzeptable Annäherung an die Version NR3.1 mit numerischen Analogskalen liefern. Zur Ähnlichkeit der Version VA3.1 mit der Version LK3.1 aus der Studie CL303 findet sich jedoch keine Aussage.

### *Zusammenfassung*

Für die eingesetzten Subskalen des WOMAC zeigen sich ausreichende Ergebnisse hinsichtlich Reliabilität und Validität. Die 3 Subskalen zeigten sich als grundsätzlich änderungssensitiv, wobei das Merkmal weniger stark für die Subskala „Physische Funktion“ ausgeprägt zu sein scheint.

Die klinischen Relevanzschwellen werden als nicht ausreichend validiert eingeschätzt. Grund dafür sind einerseits die unterschiedlichen Antwortformate der Fragebogenversion zur Ableitung und der in der Studie eingesetzten Version. Es ist unklar, ob die Bezugszeiträume der Fragebögen gleich sind. Auch der gewählte Anker der unzureichenden Response wird kritisch gesehen. Zusätzliche Unsicherheiten ergeben sich aus den großen Unterschieden in den vom pU abgeleiteten Schwellenwerten für XLH-Erkrankte. Vor diesem Hintergrund werden die Ergebnisse des WOMAC

zum Gruppenunterschied in den mittleren Veränderungen mit Hedges  $g$ , nicht jedoch die Responderanalysen dargestellt.

Zusätzlich sollte in der Interpretation beachtet werden, dass das Instrument mit einem Bezugszeitraum von 48 Stunden vor der Studienvisite und 2 Studienvisiten nach Baseline nur Aussagen für einen geringen Anteil der Studienzeit erlaubt.

### Brief Pain Inventory - Short Form (BPI-SF)

Der Endpunkt Schmerzen mittels BPI-SF wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung:

##### *Beschreibung*

Zur Erfassung von Schmerzen wurde in der Studie CL303 der BPI-SF eingesetzt.

Dabei handelt es sich um einen patientenberichteten Fragebogen mit einem Bezugszeitraum von 24 Stunden. Der Fragebogen umfasst 15 Items, von denen 11 Items quantitativ ausgewertet werden können. Der Aufbau dieser 11 Items im BPI-SF und die dahinterliegende Struktur sind in Abbildung 1 dargestellt. Der Fragebogen erfasst in 4 Items Schmerzintensität und in 7 Items Beeinträchtigung durch Schmerzen. Laut Manual zum BPI-SF wurden für die letztere Domäne 2 Subdomänen vorgeschlagen: eine affektive Subdomäne und eine Aktivitäts-Subdomäne. Die Zuteilung des Items zur Erfassung von Schlafproblemen auf diese Subdomänen ist dabei unklar. Die übrigen 4 Items des Fragebogens erfassen den Einsatz und die Wirksamkeit von Schmerzmedikation, Lokalisation der Schmerzen und einen Vergleich von Schmerzen [4]. Diese 4 Einzelitems wurden in der Studie CL303 nicht erhoben.

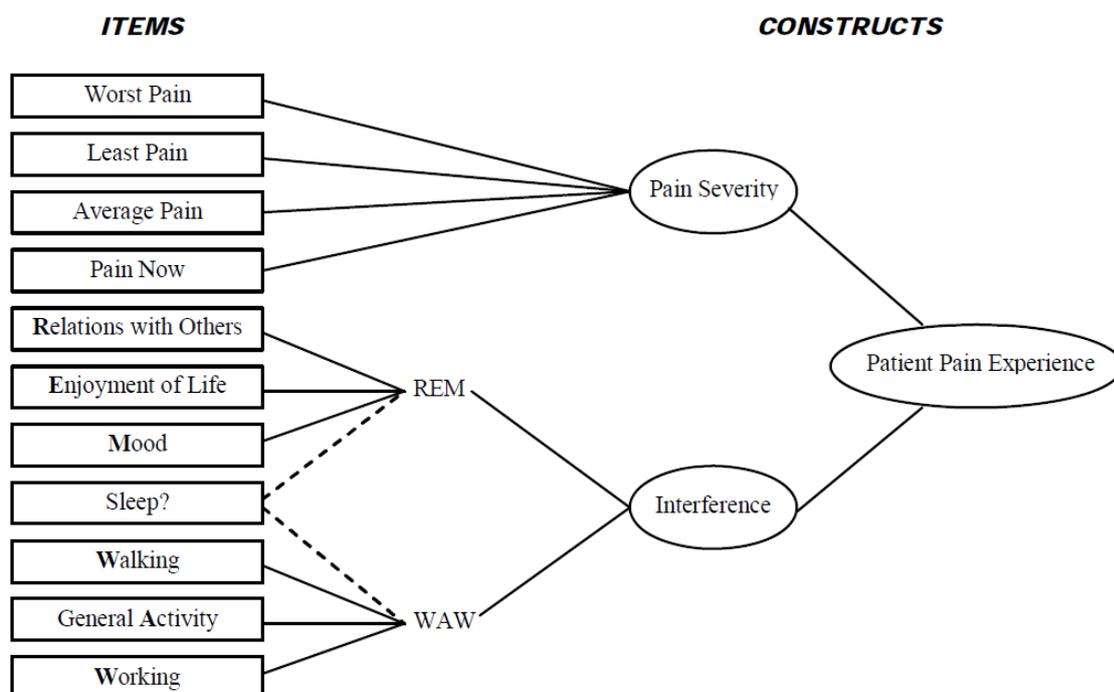


Abbildung: Aufbau und Struktur des BPI-SF. REM: relations with others, enjoyment of life, mood („affective subdimension“); WAW: walking, general activity, and work („activity subdimension“) [4]

Die Fragen der Domäne „Schmerzintensität“ werden auf einer 11-Punkte-NRS (0  $\hat{=}$  keine Schmerzen bis 10  $\hat{=}$  schlimmste vorstellbare Schmerzen) beantwortet. Für die Domäne „Schmerzbeeinträchtigung“ reicht die 11-Punkte-NRS von 0  $\hat{=}$  keine Einschränkungen bis 10  $\hat{=}$  vollständige Einschränkung. Die Domänenwerte werden als Mittelwerte der enthaltenen Items berechnet. Die Bildung eines Gesamtwerts ist nicht vorgesehen [4].

Die Erfassung des Endpunkts unterschied sich für die beiden Domänen. Die Items der Domäne „Schmerzintensität“ wurden in den 7 Tagen vor einer Studienvsiste täglich durch ein papierbasiertes Patiententagebuch erfasst. Der Studienkoordinator oder die Studienkoordinatorin rief laut Studienprotokoll die Studienteilnehmenden 8 Tage vor der Studienvsiste an und erinnerte an die Einträge. Die Items der Domäne „Schmerzintensität“ wurden zudem während der Studienvsiste erhoben. In die Auswertung ging der Mittelwert dieser Erhebungen ein. Dazu mussten Einträge für mind. 3 Tage aus der Woche vor einer Studienvsiste im Patiententagebuch vorliegen. Falls weniger Einträge verfügbar waren, wurde der Wert zur Studienvsiste imputiert. Falls auch für die Studienvsiste kein Wert vorhanden war, wurde dieser als fehlend gesetzt.

Items der Domäne „Schmerzbeeinträchtigung“ wurden lediglich während der Studienvsiste erhoben. Falls für die Studienvsiste kein Wert vorhanden war, wurde dieser als fehlend gesetzt. Während der Studienvsisten wurden PRO-Instrumente in der folgenden Reihenfolge vor Durchführung der Funktionstests beantwortet: BPI-SF, BFI, WOMAC und PGI.

Der pU hat als Wirksamkeitsendpunkt separate Auswertungen der mittleren Veränderungen in den 4 Einzelitems zur Schmerzintensität vorgesehen. Zusätzlich waren Auswertungen der mittleren Veränderungen in den Domänenwerten „Schmerzintensität“ und „Schmerzbeeinträchtigung“ a priori geplant.

Für Frage 3 zur Erfassung der schlimmsten Schmerzen der letzten 24 Stunden waren zudem Responderanalysen vorgesehen. Die folgenden Schwellenwerte werden mit einer Publikation von Dworkin et al. [8] begründet:

- Verringerung um mind. 15 % von Baseline und
- Verringerung um mind. 30 % von Baseline.
  
- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu Baseline, Woche 12 und Woche 24 erhoben. Die Items der Domäne „Schmerzintensität“ wurden in der Woche vor jeder Studienvsiste zudem über ein Patiententagebuch erfasst.
  
- Auswertung:  
Die Auswertung erfolgte auf Basis der FAS-Population mit Datenschnitt zu Woche 24.

### *Bewertung*

Die Operationalisierung des BPI-SF ist insgesamt nachvollziehbar dargestellt. Der pU begründet jedoch nicht nachvollziehbar die unterschiedlichen Erhebungsformen der Domänen „Schmerzintensität“ und „Schmerzbeeinträchtigung“ mit und ohne Patiententagebuch.

Analog zur Erhebung des WOMAC ist auch hier keine Rationale verfügbar, warum das Messinstrument nicht in der Studienvsiste bei vorzeitigem Abbruch eingesetzt wurde.

### Patientenrelevanz:

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

### Validität:

Der BPI-SF ist eine Kurzform des BPI und wurde zunächst als Wisconsin Brief Pain Questionnaire (BPQ) entwickelt [5,6]. Der BPI-SF, der ursprünglich zur Erhebung krebsbedingter Schmerzen entwickelt wurde, ist mittlerweile in einer Vielzahl klinischer Indikationen untersucht und validiert. Es existieren zahlreiche kultur- und sprachvalidierte Übersetzungen [4].

Auch für den BPI-SF hat der pU eigene Untersuchungen (siehe oben: WOMAC) durchgeführt, wobei lediglich das Item 3 zur Erfassung der schlimmsten Schmerzen außerhalb der Studie CL303 zum Einsatz kam. Die Relevanz des Items wurde in einer Interviewstudie mit 18 erwachsenen XLH-Erkrankten belegt. Anhand des Items konnte zwischen Gruppen mit und ohne Schmerzmedikation, mit und ohne Narkotika-Medikation, mit und ohne aktuellen Gebrauch einer Gehhilfe, mit und ohne Frakturen und mit leichten/moderaten und starken Knochenschmerzen unterschieden werden. Dies basiert auf Ergebnissen der Studie UX023-CL001 mit N = 201. Spearman-Korrelationen zwischen den Items zur „Schmerzintensität“ und der Domäne „Schmerzbeeinträchtigung“ zum SF-36 „Körperliche Schmerzen“ lagen zwischen -0,68 und -0,86. Die interne Konsistenz der Domäne Schmerzbeeinträchtigung lag mit Cronbachs  $\alpha > 0,81$  mind. im guten Bereich.

Es wird zudem von der Untersuchung klinischer Relevanzschwellen berichtet, denen Studie UX023-CL001 und Studie UX023-CL203 zugrunde liegen. Die eingesetzte Methodik entsprach dabei der Untersuchung von klinischen Relevanzschwellen zum WOMAC (siehe oben). Die ankerbasierte Ableitung verwendete eine Verbesserung um  $> 3$  Punkte im SF-36 „Körperliche Schmerzen“ und umfasste 13 Personen für das Item 3 des BPI-SF und 12 Personen für die Domäne „Schmerzbeeinträchtigung“. Für Item 3 entsprach dies einer prozentualen Veränderung um -23,65 % bzw. -1,62 (SD: 2,66) Punkten, für die Domäne „Schmerzbeeinträchtigung“ einer prozentualen Veränderung um -40,61 % bzw. -1,86 (SD: 1,59) Punkten. Die verteilungsbasierten Werte reichten von 0,91–1,39 für Item 3 und 0,92–1,47 für die Domäne Schmerzbeeinträchtigung [33].

Der pU führte eine Responderanalyse für Item 3 mit den Schwellenwerten einer Verringerung um mind. 15 % bzw. um mind. 30 % von Baseline durch. Bei der begründenden Referenz von Dworkin et al. handelt es sich um einen Bericht einer Konsensuskonferenz der *Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials* (IMMPACT). Hier wird von der Ableitung klinischer Relevanzschwellen von Skalen zur Erfassung der Schmerzintensität mit einer NRS von 0 bis 10 berichtet [8]. In der größten diskutierten Studie von Farrar et al. [12] wurden Daten aus 10 klinischen Studien mit insgesamt 2.724 Patientinnen und Patienten mit schmerzhafter diabetischer Neuropathie, Post-Zoster-Neuralgie, Rückenschmerzen, Fibromyalgie oder Osteoarthritis ausgewertet. Abgefragt wurden neuropathische Schmerzen der letzten 24 Stunden sowie patientenberichtete Änderungen des Gesundheitszustands. Korrelationen dieser beiden Erhebungen und die Anzahl der Befragten, die sich als verändert einstufen, sind nicht angegeben. In einer graphischen Darstellung sind über alle 10 enthaltenen Studien hinweg Verringerungen der

Schmerzintensität vor und nach der Behandlung um mind. 2 Punkte oder 30 % mit Patientenbewertungen von „stark verbessert“ verbunden. Verringerungen um mind. 4 Punkte oder 50 % wurden mit Bewertungen von „sehr stark verbessert“ in Verbindung gebracht.

In der zweiten angeführten Referenz führten Salaffi et al. [31] ROC-Analysen durch, um die absoluten und prozentualen Veränderungen zu identifizieren, die am besten zwischen den Bewertungen der Schmerzverbesserung bei 825 Patientinnen und Patienten mit Osteoarthritis, rheumatoider Arthritis oder Spondylitis ankylosans differenzierten. Die konkrete Fragestellung und der Bezugsraum der eingesetzten NRS sind nicht benannt. Nach 3 Monaten beantworteten die Patientinnen und Patienten abermals die NRS und bewerteten die Veränderung des Gesundheitszustands mittels *Patient Global Impression of Change* (PGI-C). Veränderungen in den NRS-Werten und den Kategorien des PGI-C wiesen starke Korrelationen auf (Spearman rho = 0,77 und -0,82,  $p < 0,0001$ ). Die Anzahl der Befragten, die sich als verändert einstufen, ist nicht berichtet. Verringerungen der Bewertung der Schmerzintensität um 1,0 Punkte oder 15 % im Verlauf von 3 Monaten differenzierten am besten Befragte, die ihre Schmerzen als mindestens „leicht besser“ beschrieben, von denen, die über keine Veränderung oder eine Verschlimmerung der Schmerzen berichteten. Verringerungen um 2,0 Punkte oder 33 % differenzierten am besten Befragte, die ihre Schmerzen als „viel besser“ beschrieben, von denen, die die Veränderung als nur „leicht besser“ oder „schlechter“ beschrieben.

Die dritte angeführte Publikation von Hanley et al. [16] untersuchte Relevanzschwellen für Schmerzintensität bei 34 Patientinnen und Patienten mit Amputationen und 82 Patientinnen und Patienten mit chronischen Schmerzen. Erstere bewerteten die durchschnittlichen Schmerzen in der Extremität der Amputation der letzten 24 Stunden, letztere durchschnittliche Schmerzen der letzten 24 Stunden allgemein. Verringerungen der Werte für die Schmerzintensität bei Personen um 1,8 Punkte oder 36 % entsprachen einer patientenberichteten „bedeutsamen“ Veränderung der Schmerzen (N = 33), und Verringerungen um 0,96 Punkte oder 20 % waren mit einer „spürbaren, aber nicht bedeutsamen“ Verringerung der Schmerzen (N = 31) verbunden.

In der Gesamtschau konsentiert die IMMPACT eine Veränderung um mind. 1 Punkt oder 10 % bis 20 % bei chronischen Schmerzen als geringste klinisch relevante Veränderung (MID) und eine Veränderung von mind. 2 Punkten oder 30 % als wesentliche klinisch relevante Veränderung [8].

Die Ergebnisse der dargestellten Untersuchungen scheinen trotz der angemerkten fehlenden Informationen insgesamt jedoch ausreichend robust für eine Berücksichtigung der Responderanalysen. Dabei wird der Schwellenwert der Veränderung um 15 % von Baseline im Sinne einer MID als Responseschwelle in der Nutzenbewertung herangezogen. Zur Einschätzung der Robustheit der Ergebnisse wird ergänzend die Responderanalyse mit dem Schwellenwert der Veränderung um 30 % von Baseline dargestellt.

Zusätzlich sollte in der Interpretation beachtet werden, dass das Instrument mit einem Bezugszeitraum von 24 Stunden vor der Studienvsitedate (bzw. bei Schmerzintensität die Woche vor der Studienvsitedate) und 2 Studienvsitedaten nach Baseline nur Aussagen für einen geringen Anteil der Studienzeit erlaubt.

## Brief Fatigue Inventory (BFI)

Der Endpunkt Fatigue mittels BFI wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung:

#### *Beschreibung*

Fatigue wurde in der Studie CL303 mittels BFI erhoben. Beim BFI handelt es sich um einen patientenberichteten Fragebogen zur Erfassung des Schweregrades der Fatigue und des Ausmaßes der dadurch erlebten Beeinträchtigung im täglichen Leben. Der Fragebogen umfasst 10 Items, wobei ein Item nicht quantitativ ausgewertet wird. Dieses erfragt, ob in der letzten Woche eine außergewöhnliche Fatigue vorlag. Die restlichen 9 Items sind in der folgenden Struktur aufgebaut:

- Stärke der Fatigue (3 Items)
- Beeinträchtigung durch Fatigue (6 Items).

Der Bezugszeitraum umfasst die letzten 24 Stunden. Die Fragen werden auf einer 11-Punkte-NRS (0  $\triangleq$  keine Fatigue bis 10  $\triangleq$  schlimmste vorstellbare Fatigue bzw. 0  $\triangleq$  keine Einschränkungen bis 10  $\triangleq$  vollständige Einschränkung) beantwortet. Aus den 9 Items wird als Mittelwert ein Gesamtwert berechnet, sofern mehr als 4 Items beantwortet wurden.

Die Erhebung der beiden Domänen unterschied sich analog zum BPI-SF (siehe oben). Während die Items der „Stärke der Fatigue“ in der Woche vor einer Studienvisite zusätzlich täglich über ein Patiententagebuch erfasst wurden, wurden die Items zur „Beeinträchtigung durch Fatigue“ ausschließlich während der Studienvisiten erhoben. Während der Studienvisiten wurden PRO-Instrumente in der folgenden Reihenfolge vor Durchführung der Funktionstests beantwortet: BPI-SF, BFI, WOMAC und PGI.

Der pU legt folgende Auswertungen vor:

- die Veränderung von Baseline zu jeder Visite nach Baseline in den 3 Einzelitems der Domäne „Stärke der Fatigue“, gemittelt aus den täglichen Werten des Patiententagebuchs, die über 1 Woche aufgezeichnet wurden, und den Werten der Studienvisiten sowie
- die Veränderung von Baseline zu jeder Visite nach Baseline im Gesamtwert des BFI, berechnet als Mittelwert aller 9 BFI-Items, die am Tag der Studienvisite aufgezeichnet wurden.

Dabei war die Auswertung der Einzelitems aus der Domäne „Stärke der Fatigue“ a priori lediglich deskriptiv geplant. Für das Modul 4 berechnete der pU post hoc Effektschätzer.

Zur Einschätzung der klinischen Relevanz liefert der pU zudem post hoc berechnete Hedges g.

- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu Baseline, Woche 12 und Woche 24 erhoben. Die Items der Domäne „Stärke der Fatigue“ wurden in der Woche vor jeder Studienvisite zudem über ein Patiententagebuch erfasst.
- Auswertung:  
Die Auswertung erfolgte auf Basis der FAS-Population mit Datenschnitt zu Woche 24.

### *Bewertung*

Die Operationalisierung ist insgesamt nachvollziehbar. Der pU begründet jedoch auch hier nicht nachvollziehbar die unterschiedlichen Erhebungsformen der Items aus den Domänen „Stärke der Fatigue“ und „Beeinträchtigung durch Fatigue“ mit und ohne Patiententagebuch. Der pU lässt zudem offen, ob bei der Auswertung des Gesamtwerts ebenfalls eine Berücksichtigung der Erhebungen aus dem Patiententagebuch sinnvoll gewesen wäre.

Analog zur Erhebung des WOMAC ist auch hier keine Rationale verfügbar, warum das Messinstrument nicht in der Studienvisite bei vorzeitigem Abbruch eingesetzt wurde.

### Patientenrelevanz:

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Die Erfassung von Fatigue mittels BFI wird als patientenrelevant bewertet.

### Validität:

Der BFI ist ein in seiner Validität und Reliabilität bestätigtes Messinstrument zur Erhebung von Fatigue [24]. Auch wenn nur wenige Angaben zur Entwicklung des BFI verfügbar sind, existieren zahlreiche Validierungsstudien für unterschiedliche Indikationen sowie sprach- und kulturvalidierte Übersetzungen [22,27,28,30,36,38].

Mit einer Studie zum BFI bei Patientinnen und Patienten mit rheumatoider Arthritis liegt eine Untersuchung in einer nicht-onkologischen Indikation vor [37]. 7.760 Betroffene bewerteten ihre Fatigue mittels einer Single-Item-VAS sowie mittels *Multi-dimensional Assessment of Fatigue* (MAF), Vitalitätsskala des SF-36 und BFI. Nach 6 Monaten beantworteten 5.155 Personen nochmals die Fragebögen. Korrelationen zu Baseline zwischen BFI und den anderen eingesetzten Instrumenten waren mind. moderat ausgeprägt. 13,2 % der Befragten wiesen zu Baseline den niedrigsten möglichen Wert im Gesamtwert des BFI auf, was auf einen geringen Bodeneffekt deutet. Die interne Konsistenz wurde als Cronbachs  $\alpha = 0,98$  für den BFI-Gesamtwert bestimmt und liegt damit im sehr hohen Bereich. Gleichzeitig deutet der Wert auf möglicherweise redundante Items hin. Eine Untersuchung der Änderungssensitivität wurde mittels Kendalls  $\tau$  vorgenommen. Dabei wurde die Korrelation der beobachteten Veränderungen im BFI mit Veränderungen in anderen Instrumenten überprüft. Mit  $\tau = 0,14$  zum PGI-C,  $\tau = 0,12$  zu einer Schmerz-VAS,  $\tau = 0,12$  zu einer 5-Punkt-Likert-Skala zur Gesundheitszufriedenheit und  $\tau = -0,12$  zur VAS des EuroQol zeigten sich konsistente, geringe Korrelationen. Damit erscheint der BFI grundsätzlich änderungssensitiv.

Zusätzlich sollte in der Interpretation beachtet werden, dass das Instrument mit einem Bezugszeitraum von 24 Stunden vor der Studienvisite und 2 Studienvisiten nach Baseline nur Aussagen für einen geringen Anteil der Studienzeit erlaubt.

### **Patient Global Impression of Improvement (PGI-I)**

Der Endpunkt patientenberichtete Verbesserung mittels PGI-I wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung:

#### *Beschreibung*

Der PGI-I ist eine Frage, in der anhand einer 7-Punkte-Skala (von „sehr viel besser“ bis „sehr viel schlechter“) die wahrgenommene Veränderung des Krankheitszustands seit Beginn der Studie beantwortet wird. Zu Baseline wurde die aktuelle Krankheitsschwere bezogen auf XLH über die *Patient Global Impression of Severity* (PGI-S) anhand einer 4-Punkte-Skala (von „normal“ bis „stark“) bewertet.

PRO-Instrumente wurden in der folgenden Reihenfolge vor Durchführung der Funktionstests beantwortet: BPI-SF, BFI, WOMAC und PGI.

Der pU bietet eine Auswertung als Gruppenunterschied der mittleren Veränderungen von Baseline zu Woche 24 zwischen den Studienarmen an. Zur Einschätzung der klinischen Relevanz liefert der pU zudem ein post hoc berechnetes Hedges g.

- Erhebungszeitpunkte während der Behandlungsperiode:  
Der Endpunkt wurde zu Baseline (PGI-S), Woche 12 (PGI-I) und Woche 24 (PGI-I) erhoben.
- Auswertung:  
Die Auswertung erfolgte auf Basis der FAS-Population mit Datenschnitt zu Woche 24.

#### *Bewertung*

Die Operationalisierung ist eingeschränkt nachvollziehbar. Unklar ist, welche Relevanz die Erhebung des PGI-S zu Baseline im weiteren Studienverlauf hatte.

Analog zur Erhebung des WOMAC ist auch hier keine Rationale verfügbar, warum das Messinstrument nicht in der Studienvsiste bei vorzeitigem Abbruch eingesetzt wurde.

### Patientenrelevanz:

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

Die patientenberichtete Erfassung von Veränderungen im Krankheitszustand wird als patientenrelevant erachtet.

### Validität:

In der Forschung wird die Erfassung von Veränderungen kritisch diskutiert:

Bei der rückwirkenden Einschätzung einer Veränderung müssen die Befragten in der Lage sein, sowohl ihren gegenwärtigen Zustand als auch ihren Ausgangszustand zu quantifizieren und daraus eine mentale Subtraktion durchzuführen. Dies scheint jedoch nicht möglich zu sein. Stattdessen wird bei dieser Abschätzung größtenteils der gegenwärtige Gesundheitszustand erfasst. Dieser Effekt wird umso größer, je länger das Erinnerungsintervall ist. Auch sind Erkrankungen mit hoher Symptomvariabilität besonders anfällig für diese Form der Verzerrung, dem Recall Bias [15,19,25,26]. Da es sich im vorliegenden Anwendungsgebiet jedoch um eine langsam progrediente Erkrankung handelt, erscheint dieser Faktor nur partiell übertragbar. Gleichzeitig treten etwaige Effekte durch das kontrollierte, doppelblinde Studiendesign in beiden Studienarmen auf, sodass der Endpunkt in der Nutzenbewertung berücksichtigt werden kann.

### 2.3.3 Lebensqualität

Lebensqualität wurde in Studie CL303 nicht erhoben.

### 2.3.4 Sicherheit

#### Unerwünschte Ereignisse

Der Endpunkt der unerwünschten Ereignisse wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung:

##### *Beschreibung*

Ein unerwünschtes Ereignis (UE) wurde definiert als jedes unerwünschte medizinische Ereignis, das mit dem Gebrauch des Medikaments in Verbindung stand. Ein UE konnte jedes nachteilige und unerwünschte Symptom (z. B. auch ein abnormaler Laborbefund) oder jede Erkrankung sein, die zeitlich mit der Verwendung des Arzneimittels verbunden war, unabhängig davon, ob es mit dem Arzneimittel in Zusammenhang stand oder nicht.

Es wurden klinisch signifikante Veränderungen gegenüber Baseline-Werten bei Untersuchungen, Vitalparametern, klinischen Laborparametern, Nierenultraschall, GFR, ECHO und EKG als UE erfasst, sofern dies durch die klinische Prüfperson als angemessen bewertet wurde.

Studienteilnehmende wurden bei allen Visiten nach neu aufgetretenen und weiterbestehenden UE seit der letzten Visite gefragt. Die Erhebung wurde gemäß *Medical Dictionary for Regulatory Activities* (MedDRA) v18.1. durchgeführt. Es wurden alle UE ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung bis zur letzten Studienvisite / Visite bei vorzeitigem Abbruch aufgezeichnet.

Eine anschließende telefonische Nachbeobachtung nach der letzten Studienvisite / Visite bei vorzeitigem Abbruch nach 4 und 8 Wochen fand nur bei Patientinnen und Patienten statt, die nicht unmittelbar ihre Burosumab-Therapie durch Marktzulassung oder auf anderem Wege fortsetzten.

Im SAP findet sich die Einschränkung, dass nur UE unter Behandlung (TEAE) ausgewertet werden, d. h. UE, die bei oder nach der ersten Dosis des Prüfprodukts auftraten und nicht bereits vor der ersten Dosisgabe bestanden.

#### Schwerwiegende unerwünschte Ereignisse (SUE)

Ein schwerwiegendes unerwünschtes Ereignis (SUE) ist ein UE, das nach Einschätzung des Studienarztes bzw. der Studienärztin oder des pU folgende Auswirkungen hatte:

- Tod,
- Lebensgefahr,
- Stationärer Krankenhausaufenthalt oder Verlängerung des bestehenden Krankenhausaufenthaltes,
- Anhaltende oder erhebliche Unfähigkeit oder erhebliche Beeinträchtigung der Fähigkeit, normale Lebensfunktionen auszuführen,
- Eine angeborene Anomalie/Geburtsstörung.

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

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

### Schweregrad von UE

Soweit möglich, wurde der Schweregrad der UE anhand der *Common Terminology Criteria for Adverse Events* (CTCAE v4) des *National Cancer Institute* (NCI) bewertet.

Wenn ein UE nicht nach den CTCAE-Kriterien bewertet werden konnte, sollte es als mild, mäßig, schwer, lebensbedrohlich oder tödlich unter Verwendung der folgenden Definitionen eingestuft werden:

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

### UE von besonderem Interesse

Im SAP vordefinierte Ereignisse von besonderem Interesse waren wie folgt:

- Reaktionen an der Injektionsstelle: definiert durch Preferred Terms (PT) im MedDRA-Oberbegriff (high-level term, HLT) „Reaktion an der Injektionsstelle“.
- Hypersensibilität: definiert durch relevante PT in der standardisierten MedDRA-Query (SMQ) für „Hypersensibilität“.
- Hyperphosphatämie: definiert durch PT „Hyperphosphatämie“, „Blutphosphat erhöht“ und „Blutphosphat abnormal“.
- Ektopische Mineralisierung: Da laut Studienunterlagen kein SMQ verfügbar ist, wurde das UE mit Hilfe einer MedDRA-Suche nach „Mineralisierung“ definiert.
- Syndrom der ruhelosen Beine: definiert durch die PT „Syndrom der ruhelosen Beine“, „Restlessness“, „Akathisia“, „Sinnesstörungen“, „Psychomotorische Hyperaktivität“, „Gliederschmerzen“, „Neuromuskuläre Schmerzen“ und „Formation“.
- Erhebungszeitpunkte:  
Der Endpunkt wurde zu jeder Visite erhoben. In beiden Studienarmen waren regelmäßige Hausvisiten vorgesehen. Bei Patientinnen und Patienten, die nach der letzten Visite / Visite bei vorzeitigem Abbruch nicht unmittelbar ihre Burosumab-Therapie durch Marktzulassung oder auf anderem Wege nach Studienende fortsetzten, fand eine anschließende telefonische Nachbeobachtung nach 4 und 8 Wochen statt.

- Auswertung:  
Die Auswertung erfolgte auf Basis der Safety-Population mit Datenschnitt zu Woche 24.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz:

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Die Ergebnisse zum Endpunkt unerwünschte Ereignisse werden in die Bewertung des Zusatznutzens von Burosumab eingeschlossen. Die Patientenrelevanz von Laborparametern ist unklar.

#### Validität:

In der Studie CL303 werden UE ab Unterzeichnung der Einverständniserklärung erfasst, jedoch erst ab der ersten Dosis der jeweiligen Studienmedikation ausgewertet. Es ist unklar, wieso UE erhoben, jedoch nicht ausgewertet werden.

Der Endpunkt wird bis Woche 24 dennoch als valide bewertet.

### 2.3.5 Übersicht der Erhebungszeitpunkte

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

*Tabelle 6: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie CL303*

Studienvisite (Woche) Endpunkt	Base- line	W 2	W 4	W 6	W 10	W 12	W 14	W 18	W 20	W 21	W 22	W 24	VA
Gesamtmortalität <sup>1)</sup>	x	x	x	x	x	x	x	x	x	x	x	x	x
Serumphosphat <sup>2)</sup>	x	x	x	x	x	x	x	x	x	x	x	x	x
Gehfähigkeit (6MWT)	x					x						x	x
Schmerzen, Steifheit und Physische Funktion (WOMAC)	x					x						x	
Schmerzen (BPI-SF)	x					x						x	
Fatigue (BFI)	x					x						x	
Patientenberichtete Verbesserung (PGI-I)	x <sup>3)</sup>					x						x	
Unerwünschte Ereignisse	x	x	x	x	x	x	x	x	x	x	x	x	x

<sup>1)</sup> Im Rahmen der Sicherheitserhebung erfasst.

<sup>2)</sup> Endpunkt wird ergänzend dargestellt.

<sup>3)</sup> Zu Baseline wurde der PGI-S eingesetzt.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; BFI: Brief Fatigue Inventory; BPI-SF: Brief Pain Inventory - Short Form; PGI-I: Patient Global Impression of Improvement; PGI-S: Patient Global Impression of Severity; VA: Vorzeitiger Abbruch; W: Woche; WOMAC: Western Ontario and McMaster Universities Osteoarthritis

## 2.4 Statistische Methoden

In der Studie CL303 wurde die Überlegenheit von Burosumab gegenüber Placebo im Parallelgruppendesign getestet. Der finale SAP (Version 3.0) vom 29. Oktober 2018 wurde nach dem ersten Datenschnitt zu Woche 24 (22. Dezember 2016) finalisiert.

Zuvor erschienen die Version 1 vom 28. November 2016 und die Version 2 vom 31. Januar 2017 des SAP. Wesentliche Änderungen zwischen den Versionen betreffen die folgenden Punkte:

- Änderungen von Version 2 zu Version 3 betreffen die Festlegung von Analysen für die erweiterte Extensionsperiode, die mit Amendment 3 und 4 des Studienprotokolls eingeführt wurde.
- Änderungen von Version 1 zu Version 2:
  - In der Analyse des primären und der sekundären Wirksamkeitsendpunkte wurde „Region“ als unabhängige Variable aufgenommen. Laut SAP wurde diese Änderung als Reaktion auf eine Empfehlung der *Food and Drug Administration* (FDA) umgesetzt.
  - Die Sensitivitätsanalysen mit den Imputationsmethoden LOCF (*Last Observation Carried Forward*) und BOCF (*Baseline Observation Carried Forward*) wurden ergänzt. Zuvor war lediglich eine Sensitivitätsanalyse mit mBOCF (*modified Baseline Observation Carried Forward*) vorgesehen. Auch diese Änderung wird mit der Umsetzung einer FDA-Empfehlung begründet.
  - Zusätzlich wurden Subgruppenanalysen mit den Merkmalen „Geschlecht“ und „Ethnie“ als Reaktion auf eine FDA-Empfehlung aufgenommen.

Im Studienbericht werden zudem a priori nicht geplante Subgruppenanalysen präsentiert (siehe unten: Subgruppenanalysen).

Bereits mit SAP Version 1 ist der Umgang mit einem Fehler in der stratifizierten Randomisierung beschrieben. Im Laufe der Studiendurchführung wurde festgestellt, dass das *Interactive Web Response System* (IWRS) irrtümlicherweise die Frage 5 des BPI-SF („Durchschnittliche Schmerzen“) anstelle der vorgesehenen Frage 3 („Schlimmste Schmerzen“) zur Stratifizierung der Randomisierung verwendet hatte. Dies war darauf zurückzuführen, dass die ersten beiden Items des BPI-SF nicht erhoben wurden (siehe Kapitel 2.3.2: BPI-SF) und sich dadurch die Nummerierung der Items änderte. Angesichts der Tatsache, dass die Rekrutierung der Studie zum Zeitpunkt der Feststellung des Fehlers bereits begonnen hatte, wurde beschlossen, dass die stratifizierte Randomisierung mit Frage 5 des BPI-SF fortgeführt wurde. Der IWRS-Fehler wurde als Protokollabweichung aufgezeichnet.

Zusätzlich wurden bereits verschiedene Analysen im SAP Version 1 geplant, um Auswirkungen dieses Fehlers zu untersuchen. Dazu waren die unten beschriebenen Sensitivitätsanalysen vorgesehen. Zusätzlich wurde die Stärke der Korrelation zwischen Item 3 und Item 5 des BPI-SF als Pearsons Korrelationskoeffizient bestimmt.

### Analysepopulationen

Es wurden 2 relevante Analysepopulationen in der Studie CL303 (siehe Tabelle 7) bestimmt. Die Analysen der Wirksamkeitsendpunkte wurden im Full Analysis Set, die der Sicherheitsendpunkte im Safety Set geplant.

Tabelle 7: Relevante Analysepopulationen der Studie CL303

Population	Definition
Full Analysis Set	Alle randomisierten Patientinnen und Patienten mit mindestens einer Dosis der zugeteilten Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.
Safety Set	Alle randomisierten Patientinnen und Patienten mit mindestens einer Dosis der zugeteilten Studienmedikation. Die Auswertung erfolgte entsprechend der tatsächlich erhaltenen Behandlung.

### Statistische Analysen

Statistische Tests wurden zweiseitig auf dem Signifikanzniveau  $\alpha = 0,05$  und mit 95%-Konfidenzintervallen durchgeführt.

Der primäre Endpunkt wurde mithilfe des Cochran-Mantel-Haenszel (CMH)-Tests analysiert, um die Anteile der Testpersonen in der Burosumab-Gruppe und der Placebo-Gruppe zu vergleichen, die in der Mitte des Dosiszyklus einen Serumphosphat-Wert über dem LLN erreichten. Es wurden die Stratifizierungsfaktoren der Randomisierung („Durchschnittliche Schmerzen in den letzten 7 Tagen“ [BPI-SF Item 5] und „Region“) in der Analyse berücksichtigt. Als Sensitivitätsanalyse sollte die Auswertung mit der ursprünglich geplanten Stratifizierungsvariable der Randomisierung „Schlimmste Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 3) anstelle der Stratifizierungsvariable „Durchschnittliche Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 5) durchgeführt werden. Im Modul 4 stellt der pU zusätzlich post hoc berechnete Effektschätzer dar. Die Responderanalysen zum BPI-SF wurden analog zur primären Analyse durchgeführt.

Alle Auswertungen von Unterschieden in den mittleren Veränderungen zwischen den Studienarmen wurden mittels *General Estimating Equations* (GEE)-Modell mit den Faktoren Studienarm, Visite, einem Interaktionsterm für Studienarm x Visite, Stratifizierungsfaktoren der Randomisierung („Durchschnittliche Schmerzen in den letzten 7 Tagen“ [BPI-SF Item 5] und „Region“) und einer Adjustierung für den jeweiligen Baseline-Wert durchgeführt. Als Kovarianzstruktur wurde in diesen Analysen „Compound Symmetry“ gesetzt.

Für die Auswertung der mittleren Veränderungen im BPI-SF Item 3 (Schlimmste Schmerzen) war eine Sensitivitätsanalyse vorgesehen, in der anstatt für den Baseline-Wert des Item 3 für den Baseline-Wert des Item 5 („Durchschnittliche Schmerzen“) adjustiert werden sollte. Weitere Sensitivitätsanalysen waren für WOMAC „Steifheit“ und WOMAC „Physische Funktion“ mit Stratifizierungsfaktor BPI-SF Item 3 („Schlimmste Schmerzen“) anstelle des Stratifizierungsfaktors BPI-SF Item 5 („Durchschnittliche Schmerzen“) sowie für die unten beschriebenen Imputationsregeln vorgesehen.

### Imputationsregeln

Sofern nicht anders angegeben, sollten nur beobachtete Daten verwendet werden. Wenn Veränderungen zu Baseline ausgewertet wurden, sollten nur Studienteilnehmende mit Baseline-Erhebung berücksichtigt werden. In Responder-Auswertungen wurden Personen mit fehlenden Werten als Non-Responder gewertet.

In Abhängigkeit der Anzahl der Studienabbrüche und fehlender Studienvisiten waren laut SAP für die Wirksamkeitsendpunkte Sensitivitätsanalysen möglich. Die Sensitivitätsanalysen sollten mit der LOCF-, der BOCF- und der mBOCF-Imputationsmethode durchgeführt werden. Bei der LOCF-

Analyse wurde der letzte nicht fehlende Wert nach Baseline zur Auswertung übernommen. In der BOCF-Analyse wurde der nicht fehlende Baseline-Wert zur Auswertung für den entsprechenden Endpunkt weitergeführt. Die Imputation für die mBOCF-Analyse war wie folgt definiert:

- Bei Abbruch aufgrund von UE oder Tod wurde der schlechteste Wert zwischen BOCF und LOCF verwendet;
- Andernfalls wurde LOCF verwendet.

### **Subgruppenanalysen**

Für die primären und sekundären Wirksamkeitsanalysen waren die folgenden Subgruppen a priori festgelegt:

- BPI-SF Schlimmster Schmerz zu Baseline ( $\leq 6,0$  vs.  $> 6,0$ )
- BPI-SF Durchschnittlicher Schmerz zu Baseline ( $\leq 6,0$  vs.  $> 6,0$ )
- Region (Nordamerika/EU vs. Japan vs. Südkorea)
- Geschlecht (weiblich vs. männlich)
- Ethnie (weiß vs. nicht weiß).

Laut SAP waren für die Wirksamkeitsendpunkte weitere Subgruppenanalysen möglich, wobei diese nicht weiter eingegrenzt sind. Für Sicherheitsendpunkte waren Subgruppenanalysen mit dem Merkmal „Region“ vorgesehen.

Im Studienbericht werden zusätzlich für den BPI-SF Item 3 und WOMAC „Steifheit“ und „Physische Funktion“ weitere Subgruppenanalysen, die nicht a priori geplant waren, dargestellt. Diese wurden mit den folgenden Merkmalen durchgeführt:

- Pseudofrakturen zu Baseline (ja vs. nein)
- WOMAC „Steifheit“ zu Baseline ( $\leq$  Median vs.  $>$  Median)
- WOMAC „Physische Funktion“ zu Baseline ( $\leq$  Median vs.  $>$  Median).

Im Modul 4 legt der pU Subgruppenanalysen mit den oben genannten Merkmalen für alle Endpunkte vor. Zudem werden Post-hoc-Analysen mit den folgenden Kategorien vorgelegt:

- Alter ( $\leq 41$  Jahre vs.  $> 41$  Jahre)
- Verwendung von Schmerzmitteln zu Baseline (ja vs. nein)
- Verwendung von Opioiden zu Baseline (ja vs. nein)
- WOMAC „Schmerzen“ zu Baseline ( $\leq 50$  vs.  $> 50$ )
- WOMAC „Gesamtwert“ zu Baseline ( $\leq 49,5$  vs.  $> 49,5$ )
- 6MWT zu Baseline ( $\leq 365$  m vs.  $> 365$  m).

Im Ergebnisteil werden statistisch signifikante Interaktionen ( $p < 0,05$ ) für die für die Nutzenbewertung ausgewählten Effektschätzer berichtet, sofern „[...] jede Subgruppe mindestens 10 Personen umfasst und bei binären Daten mindestens 10 Ereignisse in einer der Subgruppen aufgetreten sind“ [18].

## 2.5 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 8: Verzerrungspotential der Studie CL303 während der kontrollierten Behandlungsperiode

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung Patient/Patientin	Verblindung Studienpersonal	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
CL303	Unklar <sup>1) 2) 3)</sup>	Ja <sup>4)</sup>	Ja <sup>5)</sup>	Ja <sup>5)</sup>	Nein <sup>6)</sup>	Nein	Niedrig

<sup>1)</sup> Die Randomisierungsliste wurde laut Modul 4 erstellt durch einen Drittanbieter. Die Randomisierungsmethode ist nicht benannt.

<sup>2)</sup> Die Randomisierung war stratifiziert nach „Durchschnittliche Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 5  $\leq$  6,0 vs.  $>$  6,0) und „Region“ (Nordamerika/Europa vs. Japan vs. Südkorea). Ursprünglich war eine Stratifizierung für „Schlimmste Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 3) vorgesehen. Im IWRS wurde fälschlicherweise eine Stratifizierung für Item 5 des BPI-SF programmiert. Da die Rekrutierung bereits begonnen hatte, als der Fehler bemerkt wurde, wurde die stratifizierte Randomisierung mit Item 5 fortgeführt. Zwischen den Studienarmen zeigen sich ungleiche Verteilungen in den Kategorien des geplanten Stratifizierungsfaktors mittels Item 3 des BPI-SF: 75 % der Studienteilnehmenden aus dem Verumarm und 59 % aus dem Placebo-Arm wiesen einen Wert  $>$  6 auf. Basierend auf den Mittelwerten zeigen sich zu Baseline keine deutlichen Unterschiede zwischen den Studienarmen in diesem Merkmal (siehe Tabelle 12).

<sup>3)</sup> Hinsichtlich des Stratifizierungsfaktors „Region“ kann die Einführung mit Amendment 3 des Studienprotokolls vom 31. März 2017 (siehe Kapitel 2.2: Protokolländerungen) nicht nachvollzogen werden. Der Datenschnitt für die erste Auswertung zum Abschluss der Placebo-kontrollierten Behandlungsperiode wurde am 22. Dezember 2016 durchgeführt, sodass die Rekrutierung vor diesem Datum bereits abgeschlossen war.

<sup>4)</sup> Die Zuteilung der Randomisierung erfolgte über ein IWRS.

<sup>5)</sup> Während der Behandlungsperiode trat bei 4 Personen aus dem Burosumab-Arm eine Hyperphosphatämie (siehe Tabelle 26) auf, welche gemäß Studienprotokoll (siehe Tabelle 4) eine Entblindung und Dosisreduktion zur Folge hatte. Laut Fließtext des Studienberichts und auch laut EPAR [10] traten Hyperphosphatämien bei 5 Personen auf. Laut EPAR führten die Dosisreduktionen zudem lediglich zu einer Entblindung des Studienpersonals und nicht der betroffenen Studienteilnehmenden. Zusätzlich wird im EPAR [10] berichtet, dass es zu 2 versehentlichen Entblindungen jeweils im Burosumab-Arm kam. Aus den Studienunterlagen wird hingegen eine Entblindung als Protokollverletzung berichtet (siehe Tabelle 11). Dabei ist es auch auf Basis des EPAR [10] abschließend unklar, ob die Entblindung auch die Testperson betraf. Insgesamt betreffen diese Aspekte jedoch nur wenige Studienteilnehmende.

<sup>6)</sup> Es wurden Amendments des SAP nach Datenschnitt erstellt. Relevante Änderungen für den dargestellten Datenschnitt betreffen die Adjustierung der Analysen für die Variable „Region“. Da diese Variable ebenfalls Stratifizierungsfaktor der Randomisierung war, werden die Analysen dennoch als adäquat eingestuft und dargestellt.

Das Verzerrungspotential auf Studienebene wird als niedrig eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 9 dargestellt.

Tabelle 9: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie CL303 während der kontrollierten Behandlungsperiode

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Gesamtmortalität <sup>1)</sup>	Ja	Ja	Nein	Nein	Niedrig
Serumphosphat <sup>2)</sup>	Ja	Ja	Nein	Nein	Niedrig

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Gefähigkeit (6MWT)	Ja	Ja <sup>3)</sup>	Nein <sup>4)</sup>	Ja <sup>5) 6) 7)</sup>	Hoch
Schmerzen, Steifheit und Physische Funktion (WOMAC)	Ja	Ja <sup>3)</sup>	Nein <sup>8)</sup>	Ja <sup>7) 9) 10)</sup>	Unklar
Schmerzen (BPI-SF)	Ja	Ja <sup>3)</sup>	Nein	Ja <sup>7) 10) 11)</sup>	Unklar
Fatigue (BFI)	Ja	Ja <sup>3)</sup>	Nein	Ja <sup>6) 7) 10)</sup>	Unklar
Patientenberichtete Verbesserung (PGI-I)	Ja	Ja <sup>3)</sup>	Nein	Ja <sup>6) 7) 10)</sup>	Unklar
Unerwünschte Ereignisse	Ja	Ja	Nein <sup>12)</sup>	Ja <sup>10)</sup>	Niedrig

<sup>1)</sup> Im Rahmen der Sicherheit erfasst.

<sup>2)</sup> Ergänzend dargestellt.

<sup>3)</sup> Die Analysepopulation umfasst nicht alle randomisierten Studienteilnehmenden. Angaben zu den Gründen für die fehlenden Werte wurden nicht identifiziert. Aufgrund des geringen Anteils (< 5 %) wird das Verzerrungspotential aus diesem Aspekt jedoch dennoch als gering eingestuft.

<sup>4)</sup> Es erfolgte eine Änderung an der Datenbank nach Durchführung der a priori geplanten Auswertung im Studienbericht Woche 24. Es wurde eine erneute Auswertung für den Studienbericht Woche 48 durchgeführt. Die Änderung in der Datenbank betrifft den Baseline-Wert einer Testperson aus dem Burosumab-Arm, der aufgrund eines Berechnungsfehlers in einem Studienzentrum falsch gewesen sei. Es kann nicht nachvollzogen werden, wie der fehlerhafte Wert nach bereits erfolgter Datenbereinigung für den ursprünglichen Datenschluss entdeckt wurde und worin der Berechnungsfehler konkret bestand. In der ersten Auswertung aus Studienbericht Woche 24 zeigte sich kein statistisch signifikanter Gruppenunterschied; in der erneuten Auswertung aus Studienbericht Woche 48 hingegen schon. Da die Änderung zudem zwangsläufig nach Entblindung und unter Kenntnis der Ergebnisse aus der ersten Analyse stattfand, kann eine ergebnisgesteuerte Änderung nicht ausgeschlossen werden. In der Nutzenbewertung wird die Auswertung aus dem ersten Studienbericht Woche 24 berücksichtigt und das Verzerrungspotential für diese Auswertung bewertet.

<sup>5)</sup> Die Änderung eines Baseline-Werts für die Analysen aus den Studienberichten Woche 48 und 96 sowie der Ausschluss einer Person in der Sensitivitätsanalyse im Modul 4 des Dossiers führt zu einer Änderung in der Signifikanz des Ergebnisses. Die Robustheit der Auswertungen scheint daher eingeschränkt.

<sup>6)</sup> Es liegen keine Informationen dazu vor, ob die Annahme der Normalverteilung für die GEE-Auswertungen erfüllt war.

<sup>7)</sup> Es ist keine Sensitivitätsanalyse für den geplanten Stratifizierungsfaktor „Schlimmste Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 3) für die in der Nutzenbewertung dargestellte Auswertung verfügbar.

<sup>8)</sup> Hinsichtlich der Subskala „Steifheit“ präsentiert der pU in den Studienberichten ab Woche 48 eine erneute Auswertung anstelle der ursprünglichen Auswertung aus dem Studienbericht Woche 24. Für die neue Auswertung wurde laut Studienberichten die Datenbank nach Datenbankschluss erneut geöffnet. In den Studienunterlagen ist jedoch nicht beschrieben, welche Änderung aus welchem Grund durchgeführt wurde. Da sich die Auswertungen nur gering unterscheiden, wird nicht von einer ergebnisgesteuerten Änderung ausgegangen.

<sup>9)</sup> Die Annahme der Normalverteilung für die GEE-Auswertungen war für die Subskalen „Physische Funktion“ und „Steifheit“ laut EPAR [10] nicht erfüllt. Alternativ wurden für die EMA ANCOVA-Auswertungen vorgelegt, die nicht mit dem Dossier eingereicht wurden. Es liegen keine Informationen dazu vor, ob die Annahme der Normalverteilung für die GEE-Auswertung für die Subskala „Schmerzen“ erfüllt war.

<sup>10)</sup> Es wurden abweichende Daten und Ergebnisse für die gleichen Analysezeiträume zwischen allen Studienunterlagen festgestellt.

<sup>11)</sup> Die Annahme der Normalverteilung war für die GEE-Auswertung des Item 3 („Schlimmste Schmerzen“) laut EPAR [10] nicht erfüllt. Inwiefern dies auch auf die in der Nutzenbewertung dargestellten Domänen Schmerzintensität und Schmerzbeeinträchtigung zutrifft, ist aufgrund fehlender Informationen unklar.

<sup>12)</sup> Es werden für die Sicherheitsauswertungen aus dem Studienbericht Woche 48 Änderungen in der Datenbank nach bereits erfolgtem Datenbankschluss für den Studienbericht Woche 24 berichtet. Es wurden keine weiteren Informationen zu den vorgenommenen Änderungen identifiziert.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; ANCOVA: Kovarianzanalyse; BFI: Brief Fatigue Inventory; BPI-SF: Brief Pain Inventory - Short Form; EMA: European Medicines Agency; EPAR: European Public Assessment Report; GEE: General Estimating Equations; ITT: Intention to treat; PGI-I: Patient Global Impression of Improvement; PGI-S: Patient Global Impression of Severity; pU: pharmazeutischer Unternehmer; WOMAC: Western Ontario and McMaster Universities Osteoarthritis

### 3 Ergebnisse der eingeschlossenen Studie

#### 3.1 Studiencharakteristika und Studienmedikation

In Studie CL303 wurden insgesamt 134 Studienteilnehmende randomisiert. Dabei wurden 68 Personen dem Studienarm Burosumab und 66 Personen dem Studienarm Placebo zugeordnet (siehe Tabelle 10). Davon erhielten alle Personen mind. eine Dosis der jeweiligen Intervention und gingen somit in die Analysepopulationen *Full Analysis Set* und *Safety Set* (vgl. Tabelle 7) ein.

Alle Studienteilnehmenden mit Ausnahme einer Person aus dem Burosumab-Studienarm schlossen die 24-wöchige kontrollierte Behandlungsperiode ab. Diese Testperson, die ihre Einwilligung zur Teilnahme an der Studie zurückgezogen hatte, erhielt die letzte Dosis Burosumab etwa 1 Monat vor dem Studienabbruch und schloss die Studienvsiste zu Woche 24 nicht ab.

Es konnte keine Angabe zur medianen Beobachtungsdauer identifiziert werden. Da in der Studie CL303 während der kontrollierten Behandlungsperiode nur ein Studienabbruch festgestellt wurde, wird davon ausgegangen, dass die mediane Behandlungsdauer der medianen Beobachtungszeit entspricht.

Tabelle 10: Allgemeine Angaben zum Patientenfluss in der Studie CL303

CL303	Burosumab	Placebo
Randomisierte Patientinnen und Patienten <sup>1)</sup> , n (%)	68 (100)	66 (100)
Sicherheitspopulation, n (%)	68 (100)	66 (100)
Abbruch der Einnahme der Studienmedikation während der Behandlungsperiode, n (%)	1 (1,5)	0
Entzug der Einverständniserklärung, n (%)	1 (1,5)	0
Studienperioden		
Behandlungsperiode (Studienwoche 24) abgeschlossen, n (%)	67 (98,5)	66 (100)
Behandlungsfortsetzungsperiode (Studienwoche 48) abgeschlossen, n (%)	63 (92,6)	63 (95,5) <sup>2)</sup>
Extensionsperiode I (Studienwoche 96) abgeschlossen, n (%)	59 (86,8)	59 (89,4) <sup>2)</sup>
Extensionsperiode II <sup>3)</sup> (September 2018) abgeschlossen, n (%)	49 (72,1)	51 (77,3) <sup>2)</sup>
Studie abgeschlossen	59 (86,8)	59 (89,4) <sup>2)</sup>
Mediane Behandlungsdauer während der Behandlungsperiode, Tage (min; max)	168 (163 <sup>4)</sup> ; 177)	168 (161; 171)
Mediane Beobachtungsdauer während der Behandlungsperiode	k. A.	k. A.

<sup>1)</sup> Alle randomisierten Patientinnen und Patienten erfüllten die Bedingungen der FAS-Population (siehe Tabelle 7), sodass die FAS-Population der ITT-Population entspricht.

<sup>2)</sup> In diesen einarmigen Studienperioden erhielten alle Patientinnen und Patienten Burosumab.

<sup>3)</sup> Ein Übergang in die Extensionsperiode II war für Studienteilnehmende aus Japan und Südkorea nicht möglich.

<sup>4)</sup> Es ist unklar, ob die Angabe die Person mit Studienabbruch umfasst.

Abkürzungen: FAS: Full Analysis Set; ITT: Intention to treat; k. A.: keine Angabe; max: Maximum; min: Minimum

#### Protokollverletzungen

Protokollverletzungen werden nicht für die jeweiligen Studienperioden, sondern für die Datenschnitte berichtet. Der erste Studienbericht vom 13. Juli 2017 bezieht sich auf einen Datenschnitt vom 22. Dezember 2016, der mit Beendigung der kontrollierten Behandlungsperiode Woche 24, d. h. dem Zeitpunkt, zu dem die/der letzte Teilnehmende die Behandlungsperiode zu Woche 24 abgeschlossen hatte, durchgeführt wurde. Die medianen Expositionszeiten verlängerten

sich für die Studienarme zwischen Ende der kontrollierten Behandlungsperiode und Zeitpunkt des Datenschnitts um 81 bzw. 84 Tage. Damit ist unklar, welcher Anteil der in Tabelle 15 berichteten Protokollverletzungen während der kontrollierten Behandlungsperiode bis Woche 24 auftrat.

Bis zum Datenschnitt am 22. Dezember 2016 wiesen 60 Studienteilnehmende wesentliche Protokollverletzungen auf. Davon waren 10 Personen mehr im Burosumab-Studienarm als im Placebo-Studienarm betroffen. Die meisten wesentlichen Protokollverletzungen betrafen Prozeduren, die nicht durchgeführt wurden, andere Protokollverletzungen oder Einverständniserklärung (z. B. erneute Einverständniserklärung nicht rechtzeitig durchgeführt).

Als Beispiele für Protokollverletzungen der Kategorie „Andere“ sind im Studienbericht aufgeführt, dass Testpersonen nicht ausreichend lange nüchtern vor den Laboruntersuchungen waren bzw. dass die Dauer nicht erfasst wurde und dass eine Echokardiographie von einer Person durchgeführt wurde, die nicht im Delegationsprotokoll angegeben war.

*Tabelle 11: Protokollverletzungen in der Studie CL303 bis zum Datenschnitt 22. Dezember 2016*

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
Wesentliche Protokollverletzungen <sup>1)</sup> , n (%)	35 (51,5)	25 (37,9)
Einverständniserklärung	8 (11,8)	6 (9,1)
Prüfprodukt	2 (2,9)	1 (1,5)
Prozedur nicht durchgeführt	15 (22,1)	13 (19,7)
Einschluss- oder Ausschlusskriterien	5 (7,4)	1 (1,5)
Einnahme nicht erlaubter Begleitmedikation	1 (1,5)	0
Entblindung	1 (1,5)	0
Erhebung/Visite außerhalb des vorgesehenen Zeitfensters	1 (1,5)	1 (1,5)
Andere	13 (19,1)	11 (16,7)

<sup>1)</sup> Es konnte mehr als eine wesentliche Protokollverletzung pro Testperson vorliegen.

### **Charakterisierung der Studienpopulation**

In Tabelle 12 sind wesentliche Charakteristika der Studienpopulation dargestellt. Studienteilnehmende im Verumarm waren im Mittel 41,3 Jahre alt, Studienteilnehmende im Placebo-Arm im Mittel 38,7 Jahre alt. In beiden Studienarmen waren 35 % männliche Studienteilnehmende und 65 % weibliche Studienteilnehmende. Der überwiegende Anteil (> 85 %) der Patientinnen und Patienten nahm an der Studie über Studienzentren in Nordamerika/Europa teil.

97 (72,4 %) bzw. 37 Patientinnen und Patienten (27,6 %) wurden in die tatsächlichen Stratifizierungskategorien des BPI-SF Item 5 (Durchschnittliche Schmerzen  $\leq 6,0$  bzw.  $> 6,0$ ) aufgenommen. 44 (32,8 %) und 90 Patientinnen und Patienten (67,2 %) wurden in die geplanten Stratifizierungskategorien des BPI-SF (Schlimmste Schmerzen  $\leq 6,0$  bzw.  $> 6,0$ ) aufgenommen. Der pU argumentiert, dass bei etwa 60 % der Studienteilnehmenden die tatsächliche Stratifizierungskategorie des BPI-SF Item 5 und die geplante Stratifizierungskategorie des BPI-SF Item 3 übereinstimmten (z. B. indem die geplante Stratifizierungskategorie  $\leq 6,0$  ebenfalls der tatsächlichen Stratifizierungskategorie  $\leq 6,0$  entsprach). Der Pearson-Korrelationskoeffizient zwischen diesen Kategorien betrug 0,82256 ( $p < 0,0001$ ). Dennoch zeigen sich zwischen den Studienarmen ungleiche Verteilungen in den Kategorien des geplanten Stratifizierungsfaktors mittels Item 3 des BPI-SF: 75 % der Studienteilnehmenden aus dem Verumarm und 59 % aus dem Placebo-Arm wiesen einen Wert  $> 6$  auf. Im Mittelwert zeigen sich hingegen keine deutlichen Unterschiede zwischen den Studienarmen in der Bewertung des BPI-SF Item 3 zu Baseline.

Hinsichtlich der Anteile der Personen, die eine Schmerztherapie zu Studienbeginn und eine frühere konventionelle Therapie der XLH erhalten hatten, zeigen sich mehrheitlich gleiche Verteilungen in den Studienarmen. Im Fließtext des Studienberichts findet sich zudem die Angabe, dass insgesamt 7 der in der Studie CL303 teilnehmenden Patientinnen und Patienten zuvor in einer anderen Studie mit Burosumab behandelt wurden. Auf welche Studienarme sich diese Personen verteilen, ist nicht angegeben.

Ausgewählte Angaben zur XLH-Anamnese sind ebenfalls in Tabelle 12 zusammengefasst. Die mediane Zeit seit den ersten XLH-Symptomen wird mit 33,9 Jahren im Verumarm und 38,3 Jahren im Placebo-Arm angegeben. Die mediane Zeit seit der XLH-Diagnose wird auf 32,9 bzw. 29,7 Jahre im Verum- und Placebo-Arm geschätzt. Der pU macht allerdings keine Angaben zu den Gründen der hohen Anteile an fehlenden Werten bei diesen Erhebungen.

47 Personen im Verumarm und 38 Personen im Placebo-Arm hatten bereits Symptome einer Osteoarthritis erlebt. Nephrokalzinose und Nephrolithiasis wurden insgesamt bei 23,1 % der Patientinnen und Patienten berichtet. Hervorzuheben ist, dass in der Burosumab-Gruppe 6 Personen mehr bereits an einer Nephrokalzinose erkrankt waren als in der Placebo-Gruppe.

88 % der Testpersonen aus dem Verumarm und 83 % der Testpersonen aus dem Placebo-Arm berichteten über eine Anamnese von zahnmedizinischen/oralen Erkrankungen. Dazu gehörten häufig übermäßige Karies, Zahnextraktion, Wurzelkanalbehandlung und Zahnabszesse. Die Mehrheit der Studienteilnehmenden hatte sich bereits einer orthopädischen Operation unterzogen, wobei der Großteil mit jeweils 38 Personen pro Studienarm auf Osteotomien entfiel.

Tabelle 12: Charakterisierung der Studienpopulation der Studie CL303

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
<i>Alter (Jahre)</i>		
Mittelwert (SD)	41,3 (11,6)	38,7 (12,8)
Median (min; max)	41,9 (20,0; 63,4)	35,4 (18,5; 65,5)
<i>Altersgruppe (Jahre), n (%)</i>		
< 18	0	0
18 bis 50	52 (76,5)	56 (84,8)
51 bis 65	16 (23,5)	10 (15,2)
> 65	0	0
<i>Geschlecht, n (%)</i>		
Männlich	24 (35,3)	23 (34,8)
Weiblich	44 (64,7)	43 (65,2)
<i>Abstammung (genetisch), n (%)</i>		
Asiatisch	12 (17,6)	9 (13,6)
Schwarz oder Afro-amerikanisch	0	3 (4,5)
Kaukasisch/weiß	55 (80,9)	53 (80,3)
Andere	1 (1,5)	1 (1,5)
<i>Region <sup>1)</sup>, n (%)</i>		
Nordamerika/Europa	58 (85,3)	58 (87,9)
Japan	6 (8,8)	5 (7,6)
Südkorea	4 (5,9)	3 (4,5)

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
<i>Tatsächlicher Stratifizierungsfaktor „Durchschnittliche Schmerzen“ nach BPI-SF Item 5<sup>1)2)3)</sup>, n (%)</i>		
≤ 6,0	48 (70,6)	49 (74,2)
> 6,0	20 (29,4)	17 (25,8)
<i>Geplanter Stratifizierungsfaktor „Schlimmste Schmerzen“ nach BPI-SF Item 3<sup>2)3)</sup>, n (%)</i>		
≤ 6,0	17 (25,0)	27 (40,9)
> 6,0	51 (75,0)	39 (59,1)
<i>„Schlimmste Schmerzen“ nach BPI-SF Item 3<sup>3)</sup> zu Baseline</i>		
Mittelwert (SD)	6,8 (1,3)	6,5 (1,4)
Median (min; max)	7,1 (3,0; 8,9)	6,5 (3,1; 9,1)
<i>Serumphosphat zu Baseline (mg/dL)</i>		
Mittelwert (SD)	2,0 (0,3)	1,9 (0,3)
Median (min; max)	2,0 (1,3; 3,0)	1,9 (1,3; 2,6)
<i>Körpergröße, cm</i>	N = 67	N = 65
Mittelwert (SD)	152,2 (9,5)	152,7 (11,8)
Median (min; max)	152,0 (126,2; 176,0)	150,8 (120,6; 175,0)
<i>PHEX-Status, n (%)</i>		
Pathogen	46 (67,6)	50 (75,8)
Wahrscheinlich pathogen	8 (11,8)	7 (10,6)
Variante unklarer Bedeutung	9 (13,2)	8 (12,1)
Keine Mutation	5 (7,4)	1 (1,5)
<i>Schmerztherapie zu Studienbeginn, n (%)</i>		
Ja	47 (69,1)	44 (66,7)
Opioide	17 (25,0)	13 (19,7)
Nicht-Opioide	47 (69,1)	43 (65,2)
Nein	21 (30,9)	22 (33,3)
<i>Vortherapie mit oralem Phosphat/Vitamin D, n (%)</i>		
Nur Phosphat	3 (4,4)	1 (1,5)
Nur Vitamin-D-Metaboliten/Analoge	3 (4,4)	3 (4,5)
Beides	59 (86,8)	62 (93,9)
Keine	3 (4,4)	0
<i>Zeit seit erstmaliger Symptomatik (Jahre)</i>	N = 28	N = 32
Mittelwert (SD)	35,8 (12,4)	38,4 (13,2)
Median (min; max)	33,9 (6,0; 57,9)	38,3 (17,8; 65,6)
<i>Zeit seit XLH-Diagnose (Jahre)</i>	N = 39	N = 42
Mittelwert (SD)	31,5 (15,6)	31,4 (15,8)
Median (min; max)	32,9 (0,5; 55,8)	29,7 (0,5; 64,7)
<b>XLH-spezifische Anamnese</b>		
<i>Anamnese einer Osteoarthritis-Symptomatik, n (%)</i>		
Ja	47 (69,1)	38 (57,6)
Nein	21 (30,9)	28 (42,4)
<i>Anamnese von Knochenbrüchen, n (%)</i>		
Ja	29 (42,6)	29 (43,9)
Nein	39 (57,4)	37 (56,1)

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
<i>Anamnese von Knochenkrankungen, n (%)</i>		
Nein	4 (5,9)	3 (4,5)
Knochensporn(e)/Osteophyt(e)	29 (42,6)	25 (37,9)
Deformität der Füße	1 (1,5)	2 (3,0)
Kraniosynostose	5 (7,4)	5 (7,6)
Enthesopathie	23 (33,8)	27 (40,9)
Kniefehlstellung: Genu valgum	13 (19,1)	10 (15,2)
Kniefehlstellung: Genu varum	46 (67,6)	54 (81,8)
Innenstellung der Zehen	16 (23,5)	16 (24,2)
Rachitischer Rosenkranz	6 (8,8)	6 (9,1)
Spinalkanalstenose	15 (22,1)	10 (15,2)
Tibiatorsion	20 (29,4)	15 (22,7)
Verdickung der Handgelenke	13 (19,1)	12 (18,2)
Kniefehlstellung: Genu varum / Genu valgum	1 (1,5)	1 (1,5)
<i>Anamnese von Nierenerkrankungen, n (%)</i>		
Nein	49 (72,1)	54 (81,8)
Nephrokalzinose	11 (16,2)	5 (7,6)
Nephrolithiasis	10 (14,7)	8 (12,1)
<i>Anamnese von Zahnerkrankungen, n (%)</i>		
Nein	8 (11,8)	11 (16,7)
Zahnimplantat	9 (13,2)	12 (18,2)
Karies	42 (61,8)	32 (48,5)
Extraktionen von Erwachsenenzähnen	43 (63,2)	35 (53,0)
Wurzelkanalbehandlung	42 (61,8)	32 (48,5)
Zahnabszess	44 (64,7)	41 (62,1)
<i>Anamnese von orthopädischen Operationen, n (%)</i>		
Nein	23 (33,8)	19 (28,8)
Kraniotomie/Kraniektomie	0	1 (1,5)
Externe Fixierung	8 (11,8)	7 (10,6)
Frakturfixation mit intramedullärem Nagel/Stab	10 (14,7)	6 (9,1)
Frakturfixation mit Platten und Schrauben	11 (16,2)	8 (12,1)
Hüftprothese	4 (5,9)	3 (4,5)
Knieprothese	4 (5,9)	3 (4,5)
Osteotomie	38 (55,9)	38 (57,6)
Epiphysiodese	4 (5,9)	8 (12,1)
Korrekturchirurgie der Tibiatorsion / Deformität der Füße	3 (4,4)	3 (4,5)

<sup>1)</sup> Stratifizierungsfaktor der Randomisierung.

<sup>2)</sup> Ursprünglich war eine Stratifizierung für „Schlimmste Schmerzen“ (BPI-SF Item 3) vorgesehen. Im IWRS wurde fälschlicherweise eine Stratifizierung für „Durchschnittliche Schmerzen“ (BPI-SF Item 5) programmiert. Da die Rekrutierung bereits begonnen hatte, als der Fehler bemerkt wurde, wurde die stratifizierte Randomisierung mit Item 5 fortgeführt.

<sup>3)</sup> Höhere Werte korrespondieren mit höheren Schmerzen.

Abkürzungen: BPI-SF: Brief Pain Inventory - Short Form; cm: Centimeter; dL: Deziliter; IWRS: Interactive Web Response System; mg: Milligramm; MW: Mittelwert; PHEX: Phosphate-regulating neutral endopeptidase; SD: Standardabweichung; XLH: X-chromosomale Hypophosphatämie

### Exposition mit der Studienmedikation

Die mediane Exposition mit der Studienmedikation beträgt jeweils 168 Tage in den Studienarmen (siehe Tabelle 13).

Während der kontrollierten Behandlungsperiode erhielten die Studienteilnehmerinnen und Studienteilnehmer stabile Dosen von Burosumab. Das Studienprotokoll enthielt Kriterien für die Entblindung der Behandlungszuordnung und Dosisanpassungen aufgrund eines erhöhten Serumphosphatspiegels (siehe Tabelle 4). Bei den meisten Testpersonen waren keine Dosisanpassungen erforderlich. Laut statistischen Outputs erfüllten 6 Personen aus dem Verumarm die Kriterien für eine Dosisreduktion; im Fließtext des Studienberichts findet sich hingegen die abweichende Angabe, dass 5 Personen eine Dosisanpassung mit einhergehender Entblindung benötigten. Bei keiner Person im Placebo-Arm wurde die Behandlung entblindet. Im Fließtext des Studienberichts findet sich zudem die Angabe, dass während der kontrollierten Behandlungsperiode 2 Personen aus dem Placebo-Arm jeweils eine Dosis nicht erhielten.

*Tabelle 13: Angaben zur Exposition mit der Studienmedikation in der Studie CL303 während der kontrollierten Behandlungsperiode*

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
Mittlere Burosumab-Dosis, mg/kg (SD)		
Baseline	0,99 (0,07)	
Woche 4	0,98 (0,09)	
Woche 8	0,97 (0,11)	n. z.
Woche 12	0,96 (0,13)	
Woche 16	0,95 (0,15)	
Woche 20	0,94 (0,17)	
Woche 24	0,95 (0,18)	
Dosisreduktionen mit einhergehender Entblindung <sup>1) 2)</sup> , n (%)		
Woche 4	1 (1,5)	
Woche 8	1 (1,5)	
Woche 12	2 (3,0)	n. z.
Woche 16	1 (1,5)	
Woche 20	1 (1,5)	
Woche 24	0	

<sup>1)</sup> Laut Studienprotokoll waren Dosisreduktionen bei Überschreiten der oberen Grenze des Normalbereichs des Serumphosphats mit einhergehender Entblindung festgelegt (siehe Tabelle 4). Weitere Dosismodifikationen waren nicht vorgesehen.

<sup>2)</sup> Abweichende Angaben im Fließtext des Studienberichts, laut dessen 5 Personen Dosisreduktionen erhielten.

Abkürzungen: kg: Kilogramm; mg: Milligramm; MW: Mittelwert; n. z.: nicht zutreffend; SD: Standardabweichung

### **Begleitmedikation**

Begleitmedikation wurde in der Studie CL303 analog zu den Protokollverletzungen (siehe oben) nur für die Datenschnitte, nicht für die kontrollierte Behandlungsperiode berichtet. Daher sind die Anteile an Begleitmedikation für die kontrollierte Behandlungsperiode unbekannt.

Bis zum Datenschnitt am 22. Dezember 2016 erhielten 98,5 % der Patientinnen und Patienten aus dem Verumarm und 93,9 % der Patientinnen und Patienten aus dem Kontrollarm mind. eine Begleitmedikation (siehe Tabelle 14). Eine Differenz > 10 Prozentpunkte in den Anteilen zwischen den Studienarmen zeigte sich für Analgetika, Antiphlogistika und Antirheumatika, Vitamine sowie Husten- und Erkältungsmittel. Dabei betrafen die höheren Anteile jeweils den Verumarm.

Ebenfalls hier dargestellt ist die Erhebung der Schmerzmittel aus dem Patiententagebuch, welches in der Woche vor der Studienvsiste zum Einsatz kam. Zu Woche 24 hatten 45 Personen aus dem

Burosumab-Arm und 40 Personen aus dem Placebo-Arm die Einnahme von Schmerzmitteln in der Woche vor Studienvisite berichtet.

Tabelle 14: Begleitmedikation > 20 % in einem Studienarm in der Studie CL303

ATC-Klassifikationsgruppe 2 Preferred Term	Burosumab N = 68 n (%)	Placebo N = 66 n (%)
<b>Begleitmedikation bis zum Datenschnitt 22. Dezember 2016</b>		
Personen mit mind. einer Begleitmedikation	67 (98,5)	62 (93,9)
Analgetika	53 (77,9)	45 (68,2)
Paracetamol	35 (51,5)	28 (42,4)
Antiphlogistika und Antirheumatika	49 (72,1)	40 (60,6)
Ibuprofen	27 (39,7)	21 (31,8)
Antibiotika zur systemischen Anwendung	25 (36,8)	30 (45,5)
Vitamine	29 (42,6)	21 (31,8)
Colecalciferol	21 (30,9)	17 (25,8)
Alle übrigen therapeutischen Mittel	16 (23,5)	19 (28,8)
Alle übrigen therapeutischen Mittel	16 (23,5)	19 (28,8)
Antihistaminika zur systemischen Anwendung	20 (29,4)	15 (22,7)
Mittel bei säurebedingten Erkrankungen	17 (25,0)	18 (27,3)
Psychoanaleptika	14 (20,6)	16 (24,2)
Sexualhormone und Modulatoren des Genitalsystems	15 (22,1)	14 (21,2)
Husten- und Erkältungsmittel	17 (25,0)	7 (10,6)
<b>Schmerzmittel zu Woche 24 <sup>1)</sup></b>		
Personen mit mind. einem Schmerzmittel	45 (66,2)	40 (60,6)
Opioide	16 (23,5)	14 (21,2)
Nicht-Opioide Schmerzmittel	43 (63,2)	38 (57,6)
Neuropathische Schmerzmedikamente/Antidepressiva	4 (5,9)	3 (4,5)
Andere Schmerzmittel	5 (7,4)	2 (3,0)

<sup>1)</sup> Die Verwendung von Schmerzmitteln basiert auf den Aufzeichnungen aus den Patiententagebüchern, die in den 7 Tagen vor Studienvisite ausgefüllt wurden. Nicht-Opioide Schmerzmittel sind definiert als nicht-steroidale Antirheumatika und Paracetamol/Paraminophen.

Abkürzungen: ATC: Anatomisch-Therapeutisch-Chemisch

## 3.2 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserhebung erfasst. Im Laufe der kontrollierten Behandlungsperiode sind in der Studie CL303 keine Todesfälle aufgetreten.

## 3.3 Morbidität

### Serumphosphat

In der Studie CL303 wurden die Anteile in den Behandlungsgruppen ausgewertet, die einen mittleren Serumphosphatwert oberhalb der unteren Grenze des Normalbereichs in der Mitte und zum Ende eines Dosiszyklus aufwiesen. Dabei fand die Auswertung des Erreichens zum Ende des Dosiszyklus ausschließlich deskriptiv statt: 46 Personen aus dem Burosumab-Arm (67,6 %) und 4 Personen aus dem Placebo-Arm (6,1 %) erreichten einen mittleren Serumphosphatspiegel  $\geq 2,5$  mg/dL am Ende des Dosiszyklus.

Die Auswertung des Erreichens in der Mitte des Dosiszyklus ist in Tabelle 15 als ergänzend berücksichtigter Endpunkt dargestellt. A priori war ein Cochran-Mantel-Haenszel-Test festgelegt, der einen statistisch signifikanten Unterschied zwischen den Studienarmen zeigt. Der pU liefert in seinem Dossier ein post hoc berechnetes Relatives Risiko, für welches jedoch keine Informationen zur Teststatistik und möglichen Adjustierung identifiziert wurden. Die Ergebnisse werden daher lediglich im Anhang (siehe Tabelle 28) berichtet.

Zur Dauer der Response wurden keine Angaben identifiziert.

Sofern eine Testperson die obere Grenze des Normalbereichs im Serumphosphatspiegel überschritt, war laut Studienprotokoll die Reduktion der Burosumab-Dosis mit einhergehender Entblindung vorgesehen. Dies war bei 5 bzw. 6 Personen aus dem Burosumab-Arm während der kontrollierten Behandlungsperiode notwendig (siehe Kapitel 3.1). In der vom pU gewählten Responder-Definition sind diese Personen als Responder umfasst.

*Tabelle 15: Personen mit mittlerem Serumphosphatspiegel  $\geq 2,5$  mg/dL in der Mitte des Dosiszyklus während der Behandlungsperiode, Full Analysis Set <sup>1)</sup>*

CL303	Burosumab N = 68	Placebo N = 66
Personen mit mittlerem Serumphosphatspiegel $\geq 2,5$ mg/dL in der Mitte des Dosiszyklus n (%)	63 (92,6)	5 (7,6)
Effektschätzer [95%-KI]; p-Wert <sup>2)</sup>	n. v. [n. v.]; < 0,0001	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> A priori festgelegter Cochran-Mantel-Haenszel-Test adjustiert für den tatsächlichen Stratifizierungsfaktor des BPI-SF 5 („Durchschnittliche Schmerzen“) und Region. Die Berechnung eines zugehörigen Effektschätzers war a priori nicht geplant. Im Anhang wird ein post hoc berechnetes Relatives Risiko berichtet (siehe Tabelle 28).

Abkürzungen: dL: Deziliter; KI: Konfidenzintervall; mg: Milligramm; n. v.: nicht verfügbar

### 6-Minuten-Gehtest (6MWT)

In der Studie CL303 wurde der 6MWT zur Erfassung der körperlichen Funktion eingesetzt. Die Ergebnisse aus der Erhebung zu Baseline, den Veränderungen bis Woche 24 und der

Gruppenunterschied in den Veränderungen sind in Tabelle 16 dargestellt. Es zeigt sich kein statistisch signifikanter Unterschied zwischen den Studienarmen zu Woche 24.

Dargestellt sind im Folgenden die Ergebnisse aus dem ersten Studienbericht mit Ergebnissen aus Woche 24. In den folgenden Studienberichten zu Woche 48 und Woche 96 stellt der pU eine erneute Auswertung der Ergebnisse der Behandlungsperiode bis Woche 24 dar. In dieser Auswertung wurde laut Studienberichten die Datenbank nach Datenbankschluss erneut geöffnet, um einen fehlerhaften Baseline-Wert zu korrigieren. Dieser sei auf einen Berechnungsfehler im Studienzentrum zurückzuführen. In der Konsequenz verändert sich das Maximum im Burosumab-Studienarm von 860 Meter auf 643 Meter. Unklar ist, wie der fehlerhafte Wert nach bereits erfolgter Datenbereinigung für den ursprünglichen Datenschluss entdeckt wurde und worin der Berechnungsfehler konkret bestand. Da die Änderung zudem zwangsläufig nach Entblindung und unter Kenntnis der Ergebnisse aus der ersten Analyse stattfand, werden die Ergebnisse der ersten Analyse präsentiert.

Tabelle 16: Veränderung im 6MWT zu Woche 24, Full Analysis Set <sup>1)</sup>

CL303	Burosumab N = 68	Placebo N = 66
6MWT-Distanz zu Baseline (Meter) MW (SD)	N = 67 362,8 (125,0)	N = 65 367,4 (103,4)
Veränderung der 6MWT-Distanz zu Woche 24 (Meter) LS Mean (SE) <sup>2)</sup>	N = 65 5,9 (10,7)	N = 64 -5,7 (8,2)
LS Mean-Differenz [95%-KI]; p-Wert <sup>2)</sup>	11,6 [-8,9; 32,2]; 0,2671	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen und dem Baseline-Wert im 6MWT als Kovariate.

Abkürzungen: 6MWT: 6-Minuten-Gehtest; GEE: Generalized estimation equation; KI: Konfidenzintervall; LS: Least Squares; SD: Standardabweichung; SE: Standardfehler

### Sensitivitäts- und Subgruppenanalysen

In seinem Dossier präsentiert der pU eine post hoc durchgeführte Sensitivitätsanalyse, in der die Person mit dem betreffenden Baseline-Wert von der Analyse ausgeschlossen wurde. Mit einer LS Mean-Differenz von 21,1 (95%-KI: 5,5; 36,7) zeigte sich hier ein statistisch signifikanter Unterschied ( $p = 0,0080$ ) zwischen den Studienarmen zugunsten von Burosumab. Das ebenfalls post hoc berechnete Hedges g betrug 0,5 (95%-KI: 0,1; 0,8).

Andere Sensitivitätsanalysen wurden für diesen explorativen Endpunkt nicht geplant und durchgeführt.

Interaktionstests zur Auswertung aus dem ersten Studienbericht wurden nicht vorgelegt.

### **Western Ontario and McMaster Universities Osteoarthritis (WOMAC)-Fragebogen**

Mit dem WOMAC-Fragebogen wurden patientenberichtete Bewertungen von Schmerzen, Steifheit und physischer Funktion erfasst. Die Ergebnisse sind in Tabelle 17 präsentiert.

Für die Subskala „Schmerzen“ war a priori keine Auswertung vorgesehen. In den Studienberichten werden die Ergebnisse dieser Subskala deskriptiv dargestellt. Der pU wertete die Subskala „Schmerzen“ darüber hinaus post hoc mit einem GEE-Modell analog zu den anderen Subskalen aus. Unklar ist, woher Unterschiede in den Baseline-Werten für den Placebo-Arm in den

statistischen Outputs aus den Studienberichten und dem Nachberechnungsdokument rühren. Im Studienbericht Woche 24 wird der Mittelwert zu Baseline im Placebo-Arm mit 48,11 (SD: 15,66) angegeben, im Studienbericht Woche 96 und im Nachberechnungsdokument hingegen mit 47,95 (SD: 15,54). Für den Burosumab-Arm finden sich keine Abweichungen. In Tabelle 17 ist der Baseline-Wert aus dem Studienbericht Woche 96 bzw. Nachberechnungsdokument dargestellt, da dieser der Post-hoc-Analyse zugrunde liegt.

Hinsichtlich der Subskala „Steifheit“ präsentiert der pU in den Studienberichten ab Woche 48 eine andere Auswertung als die hier dargestellte Auswertung aus dem Studienbericht Woche 24. Für die Auswertung in den Studienberichten ab Woche 48 wurde die Datenbank nach Datenbankschluss erneut geöffnet. Jedoch ist in den Studienunterlagen nicht beschrieben, welche Änderung aus welchem Grund durchgeführt wurde. Da die Änderung nach Entblindung und unter Kenntnis der Ergebnisse aus der ersten Auswertung geschah, werden die Ergebnisse der ursprünglichen Auswertung aus dem ersten Studienbericht Woche 24 für die Nutzenbewertung herangezogen. Da für die Post-hoc-Berechnung des WOMAC-Gesamtwerts jedoch die Analyse aus dem Studienbericht Woche 48 verwendet wurde, kann dieser nicht berücksichtigt werden.

In den Ergebnissen zeigt sich kein statistisch signifikanter Unterschied zwischen den Studienarmen in den mittleren Veränderungen von Baseline bis Woche 24 für die Subskala „Schmerzen“. Für die Subskalen „Steifheit“ und „Physische Funktion“ zeigen sich hingegen statistisch signifikante Unterschiede zugunsten von Burosumab.

Die klinische Relevanz des Gruppenunterschieds in der Subskala „Steifheit“ kann nicht beurteilt werden, da Hedges g nur für die erneute Auswertung aus dem Studienbericht Woche 48 vorgelegt wurde. Für die Subskala „Physische Funktion“ liegt das 95%-Konfidenzintervall des Hedges g nicht vollständig außerhalb von  $[0,2]$ , sodass die klinische Relevanz des Effekts unklar ist.

Tabelle 17: Veränderung im WOMAC zu Woche 24, Full Analysis Set <sup>1)</sup>

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
<b>Schmerzen <sup>2) 3)</sup></b>		
Wert zu Baseline MW (SD)	N = 67 50,7 (18,0)	48,0 (15,5) <sup>4)</sup>
Veränderung zu Woche 24 LS Mean (SE) <sup>5)</sup>	N = 66 -7,1 (2,5)	N = 65 -2,8 (2,8)
LS Mean-Differenz [95%-KI]; p-Wert <sup>5)</sup>	-4,4 [-9,3; 0,6]; 0,0848	
<b>Steifheit <sup>6)</sup></b>		
Wert zu Baseline MW (SD)	64,7 (20,3)	61,7 (21,2)
Veränderung zu Woche 24 LS Mean (SE) <sup>5)</sup>	N = 67 -7,9 (3,0)	N = 65 0,3 (3,1)
LS Mean-Differenz [95%-KI]; p-Wert <sup>5)</sup>	-8,1 [-14,5; -1,8]; 0,0122	
Hedges g [95%-KI]	k. A.	

CL303	Burosumab N = 68	Placebo N = 66
<b>Physische Funktion <sup>7)</sup></b>		
Wert zu Baseline MW (SD)	50,8 (19,7)	43,9 (19,9)
Veränderung zu Woche 24 LS Mean (SE) <sup>5)</sup>	N = 66 -3,1 (2,6)	N = 65 1,8 (2,7)
LS Mean-Differenz [95%-KI]; p-Wert <sup>5)</sup>	-4,9 [-9,8; -0,1]; 0,0478	
Hedges g [95%-KI]	-0,4 [-0,7; 0,0]	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> Höhere Werte korrespondieren mit höheren Schmerzen.

<sup>3)</sup> Post hoc ausgewertet für Modul 4 des Dossiers. Die Darstellung erfolgte zuvor ausschließlich deskriptiv.

<sup>4)</sup> Angabe aus Studienbericht Woche 96 / Nachberechnungsdokument. Abweichende Angabe im Studienbericht Woche 24 mit 48,1 (SD: 15,7).

<sup>5)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen und dem jeweiligen WOMAC-Subskalenwert zu Baseline als Kovariate.

<sup>6)</sup> Höhere Werte korrespondieren mit stärkerer Steifheit.

<sup>7)</sup> Höhere Werte korrespondieren mit eingeschränkterer körperlicher Funktion.

Abkürzungen: GEE: Generalized estimation equation; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; SD: Standardabweichung; SE: Standardfehler; WOMAC: Western Ontario and McMaster Universities Osteoarthritis

### Sensitivitäts- und Subgruppenanalysen

Die Sensitivitätsanalysen für die Subskala „Physische Funktion“ und die Subskala „Steifheit“ mit Adjustierung für den geplanten Stratifizierungsfaktor des BPI-SF Item 3 („Schlimmste Schmerzen“) zeigten vergleichbare Ergebnisse hinsichtlich Stärke des Effekts und Signifikanz. Auch die Sensitivitätsanalysen mit den Imputationsmethoden LOCF, BOCF und mBOCF lieferten ähnliche Resultate wie die in Tabelle 17 dargestellte Analyse.

Analoge Sensitivitätsanalysen für die post hoc ausgewertete Subskala „Schmerzen“ liegen nicht vor.

Interaktionstests zu den Subgruppenmerkmalen zur Auswertung der Subskala „Steifheit“ aus dem ersten Studienbericht Woche 24 wurden nicht vorgelegt. Für die Subskalen „Physische Funktion“ und „Schmerzen“ zeigten sich keine Belege für eine Effektmodifikation für Analysen, in denen alle Subgruppen mindestens 10 Personen umfassten.

### **Brief Pain Inventory - Short Form (BPI-SF)**

Schmerzen wurden in der Studie CL303 ebenfalls über den BPI-SF erhoben. Anhand des Items 3 „Schlimmste Schmerzen“ wurden Responderanalysen für Verringerungen um  $> 15$  und  $> 30$  % durchgeführt. Dabei wurden Personen mit fehlendem Wert oder Studienabbruch als Non-Responder gezählt. Die Ergebnisse sind in Tabelle 18 abgebildet, wobei die Auswertung mit dem Schwellenwert der Veränderung um 30 % von Baseline ergänzend dargestellt wird (siehe Kapitel 2.3.2: BPI-SF).

Untersuchungen zum Anteil der Personen mit Verschlechterungen wurden nicht durchgeführt.

Da keine Angaben zur Teststatistik und möglichen Adjustierung der post hoc berechneten Relativen Risiken identifiziert wurden, werden diese lediglich im Anhang (siehe Tabelle 29) berichtet.

Tabelle 18: Personen mit Response im BPI-SF Item 3 zu Woche 24, Full Analysis Set <sup>1)</sup>

CL303	Burosumab N = 68	Placebo N = 66
Personen mit Verringerung > 15 % von Baseline zu Woche 24 im BPI-SF Item 3 <sup>2)</sup>		
n (%)	29 (42,6)	23 (34,8)
Effektschätzer [95%-KI]; p-Wert <sup>3)</sup>	n. v. [n. v.]; 0,3564	
Personen mit Verringerung > 30 % von Baseline zu Woche 24 im BPI-SF Item 3 <sup>2) 4)</sup>		
n (%)	17 (25,0)	12 (18,2)
Effektschätzer [95%-KI]; p-Wert <sup>3)</sup>	n. v. [n. v.]; 0,2858	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> Höhere Werte korrespondieren mit höheren Schmerzen.

<sup>3)</sup> A priori festgelegter Cochran-Mantel-Haenszel-Test adjustiert für den tatsächlichen Stratifizierungsfaktor des BPI-SF 5 („Durchschnittliche Schmerzen“) und Region. Die Berechnung eines zugehörigen Effektschätzers war a priori nicht geplant. Im Anhang wird ein post hoc berechnetes Relatives Risiko berichtet (siehe Tabelle 29).

<sup>4)</sup> Ergänzend dargestellt.

Abkürzungen: BPI-SF: Brief Pain Inventory - Short Form; KI: Konfidenzintervall; n. v.: nicht verfügbar

Die Veränderungen in den Domänenwerten sind in Tabelle 19 abgebildet. Es zeigen sich keine statistisch signifikanten Unterschiede in den Veränderungen von Baseline zu Woche 24 zwischen den Studienarmen. Die dargestellten Ergebnisse stammen aus dem Studienbericht Woche 24. Abweichende Angaben finden sich im Vergleich zum Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument, in welchem der pU erneute Auswertungen für das Modul 4 präsentiert. Laut statistischen Outputs liegt jeweils ein gleich aufgebautes GEE-Modell den Berechnungen zugrunde.

Tabelle 19: Veränderung im BPI-SF zu Woche 24, Full Analysis Set <sup>1)</sup>

CL303	Burosumab N = 68	Placebo N = 66
<b>Schmerzintensität <sup>2)</sup></b>		
Wert zu Baseline	N = 68	N = 66
MW (SD)	5,2 (1,5)	4,9 (1,5)
Veränderung zu Woche 24	N = 67	N = 65
LS Mean (SE) <sup>3)</sup>	-0,5 (0,2) <sup>4)</sup>	-0,1 (0,2) <sup>5)</sup>
LS Mean-Differenz [95%-KI]; p-Wert <sup>3)</sup>	-0,4 [-0,9; 0,1]; 0,0844 <sup>6)</sup>	
<b>Schmerzbeeinträchtigung <sup>7)</sup></b>		
Wert zu Baseline	N = 68	N = 66
MW (SD)	5,2 (2,2)	4,8 (2,2) <sup>8)</sup>
Veränderung zu Woche 24	N = 67	N = 65
LS Mean (SE) <sup>3)</sup>	-0,4 (0,2)	-0,3 (0,2) <sup>9)</sup>
LS Mean-Differenz [95%-KI]; p-Wert <sup>3)</sup>	-0,2 [-0,7; 0,4]; 0,5476 <sup>10)</sup>	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> Höhere Werte korrespondieren mit höherer Schmerzintensität.

<sup>3)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen und dem jeweiligen BPI-SF-Domänenwert zu Baseline als Kovariate.

<sup>4)</sup> Angabe aus Studienbericht Woche 24. Gleiche Angabe im Studienbericht Woche 96. Abweichende Angabe im Studienbericht Woche 48 und Nachberechnungsdokument -0,6 (SE: 0,2).

- <sup>5)</sup> Angabe aus Studienbericht Woche 24. Gleiche Angabe im Studienbericht Woche 96. Abweichende Angabe im Studienbericht Woche 48 und Nachberechnungsdokument: -0,2 (SE: 0,2).
- <sup>6)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48: 0,4 (95%-KI: -0,9; 0,1); p = 0,0926. Keine Berechnung einer LS Mean-Differenz im Studienbericht Woche 96. Wiederum abweichende Angabe im Nachberechnungsdokument: -0,4 (95%-KI: -0,9; 0,1); p = 0,0940.
- <sup>7)</sup> Höhere Werte korrespondieren mit höherer Schmerzbeeinträchtigung.
- <sup>8)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96 und im Nachberechnungsdokument: 4,7 (SD: 2,1).
- <sup>9)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und Nachberechnungsdokument: -0,3 (SE: 0,3).
- <sup>10)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48: -0,1 (95%-KI: -0,7; 0,4); p = 0,6511. Keine Berechnung einer LS Mean-Differenz im Studienbericht Woche 96. Wiederum abweichende Angabe im Nachberechnungsdokument: -0,1 (95%-KI: -0,7; 0,4); p = 0,6527.

Abkürzungen: BPI-SF: Brief Pain Inventory - Short Form; GEE: Generalized estimation equation; KI: Konfidenzintervall; LS: Least Squares; SD: Standardabweichung; SE: Standardfehler

### Sensitivitäts- und Subgruppenanalysen

Es wurden keine Sensitivitätsanalysen für die Responderanalysen sowie die Domänen „Schmerzintensität“ und „Schmerzbeeinträchtigung“ geplant und durchgeführt.

Aufgrund der Unterschiede zwischen Nachberechnungsdokument und Analysen aus dem Studienbericht Woche 24 können die vorgelegten Interaktionstests aus dem Nachberechnungsdokument nicht berücksichtigt werden.

### **Brief Fatigue Inventory (BFI)**

Fatigue wurde in der Studie CL303 durch den BFI erhoben. Die Ergebnisse sind in Tabelle 20 dargestellt. Es zeigt sich kein statistisch signifikanter Unterschied zwischen den Studienarmen in den Veränderungen von Baseline bis Woche 24.

Auch hier finden sich unterschiedliche Angaben im Studienbericht Woche 24 und dem Studienbericht Woche 48, Woche 96 bzw. dem Nachberechnungsdokument. Laut statistischen Outputs liegt jeweils ein analoges GEE-Modell den Berechnungen zugrunde.

*Tabelle 20: Veränderung im BFI <sup>1)</sup> zu Woche 24, Full Analysis Set <sup>2)</sup>*

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
<i>Wert zu Baseline MW (SD)</i>	5,4 (2,0)	4,9 (1,9)
<i>Veränderung zu Woche 24 LS Mean (SE) <sup>3)</sup></i>	N = 67 0,0 (0,3) <sup>4)</sup>	N = 65 0,0 (0,3) <sup>5)</sup>
<i>LS Mean-Differenz [95%-KI]; p-Wert <sup>3)</sup></i>	0,1 [-0,5; 0,6]; 0,7912 <sup>6)</sup>	

<sup>1)</sup> Höhere Werte korrespondieren mit stärkerer Fatigue.

<sup>2)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>3)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen und dem BFI-Gesamtwert zu Baseline als Kovariate.

<sup>4)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96/ Nachberechnungsdokument: 0,1 (SE: 0,3).

<sup>5)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 / Studienbericht Woche 96 / Nachberechnungsdokument: -0,1 (SE: 0,3).

<sup>6)</sup> Angabe aus Studienbericht Woche 24. Keine Berechnung einer LS Mean-Differenz im Studienbericht Woche 96. Abweichende Angabe im Nachberechnungsdokument: -0,1 (95%-KI: -0,7; 0,4); p = 0,6527. Ebenfalls abweichende Angabe im Studienbericht Woche 48: 0,1 (-0,5; 0,7), p = 0,7129

Abkürzungen: BFI: Brief Fatigue Inventory; GEE: Generalized estimation equation; KI: Konfidenzintervall; LS: Least Squares; SD: Standardabweichung; SE: Standardfehler

### Sensitivitäts- und Subgruppenanalysen

Es wurden keine Sensitivitätsanalysen für den Gesamtwert des BFI durchgeführt.

Aufgrund der Unterschiede zwischen Nachberechnungsdokument und Analysen aus dem Studienbericht Woche 24 können die vorgelegten Interaktionstests aus dem Nachberechnungsdokument nicht berücksichtigt werden.

### **Patient Global Impression of Improvement (PGI-I)**

Die patientenberichtete Bewertung des Krankheitszustands erfolgte in der Studie CL303 zu Baseline über PGI-S und im weiteren Studienverlauf durch PGI-I. Der Unterschied zwischen den Studienarmen in den Veränderungen zu Woche 24 ist nicht statistisch signifikant (siehe Tabelle 21).

Auch bei diesem Endpunkt finden sich unterschiedliche Angaben in den Studienunterlagen. Hier weicht jedoch auch die Modellbeschreibung ab: während in der Analyse zu Woche 24 der Stratifizierungsfaktor „Region“ im Modell berücksichtigt wurde, wird der Faktor in den späteren Auswertungen nicht berücksichtigt.

*Tabelle 21: Veränderung im PGI-I zu Woche 24, Full Analysis Set <sup>1)</sup>*

<b>CL303</b>	<b>Burosumab N = 68</b>	<b>Placebo N = 66</b>
<i>Wert zu Baseline <sup>2)</sup></i> MW (SD)	3,2 (0,6)	3,0 (0,8)
<i>Veränderung zu Woche 24</i> LS Mean (SE) <sup>3) 4)</sup>	N = 67 3,6 (0,2)	N = 64 3,9 (0,2)
LS Mean-Differenz [95%-KI]; p-Wert <sup>4)</sup>	-0,3 [-0,7; 0,2]; 0,2035 <sup>5)</sup>	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> Zu Baseline wurde PGI-S mit einer 4-Punkte-Skala eingesetzt. Höhere Werte korrespondieren mit stärkerer Krankheitslast.

<sup>3)</sup> Bei allen Studienvisiten nach Baseline wurde der PGI-I mit einer 7-Punkte-Skala eingesetzt. Höhere Werte korrespondieren mit Verschlechterungen.

<sup>4)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen.

<sup>5)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: -0,3 (95%-KI: -0,7; 0,2); p = 0,2161. Keine Berechnung einer LS Mean-Differenz im Studienbericht Woche 96.

Abkürzungen: GEE: Generalized estimation equation; KI: Konfidenzintervall; LS: Least Squares; PGI-I: Patient Global Impression of Improvement; SD: Standardabweichung; SE: Standardfehler

### Sensitivitäts- und Subgruppenanalysen

Es wurden keine Sensitivitätsanalysen für den PGI-I durchgeführt.

Aufgrund der Unterschiede zwischen Nachberechnungsdokument und Analysen aus den Studienbericht Woche 24 können die vorgelegten Interaktionstests aus dem Nachberechnungsdokument nicht berücksichtigt werden.

### 3.4 Lebensqualität

Lebensqualität wurde in der Studie CL303 nicht erfasst.

### 3.5 Sicherheit

Im folgenden Kapitel sind Ergebnisse der Sicherheitsauswertungen für die kontrollierte Behandlungsperiode bis Woche 24 der Studie CL303 dargestellt.

Unerwünschte Ereignisse (UE) wurden ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung bis zur letzten Studienvsiste / Visite bei vorzeitigem Abbruch aufgezeichnet. Ausgewertet wurden UE, die bei oder nach der ersten Dosis des Prüfprodukts auftraten und nicht bereits vor der ersten Dosisgabe bestanden (siehe Kapitel 2.3.4). Die mediane Behandlungsdauer beträgt jeweils 168 Tage in den Studienarmen (siehe Tabelle 10).

Es ist anzumerken, dass einige der im Folgenden dargestellten SOC und PT XLH-Symptome umfassen und daher nicht auszuschließen ist, dass Symptome aus der Endpunktkategorie „Morbidity“ enthalten sind.

Die zusammenfassende Darstellung der UE ist in Tabelle 22 abgebildet. Laut Studienbericht Woche 24 trat im Kontrollarm ein Ereignis mehr mit Schweregrad 3/4 als im Interventionsarm auf. Im Studienbericht Woche 96 und Nachberechnungsdokument für das Dossier werden hingegen jeweils 8 betroffene Personen für beide Studienarme berichtet. Aufgrund dieser Diskrepanz kann das post hoc berechnete Relative Risiko nicht berücksichtigt werden. Darüber hinaus liegen keine Informationen zur Teststatistik und möglichen Adjustierung zu den Relativen Risiken aus dem Nachberechnungsdokument vor. Sowohl im Interventionsarm als auch im Kontrollarm traten jeweils 2 schwerwiegende unerwünschte Ereignisse (SUE) auf. Auch hier findet sich eine Diskrepanz in den Studienunterlagen, sodass das post hoc berechnete Relative Risiko nicht herangezogen werden kann. Es führte kein UE zum Abbruch der Studie oder Studienmedikation.

*Tabelle 22: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsperiode im Safety Set <sup>1)</sup>*

Personen mit mindestens einem ...	Burosumab N = 68 n (%)	Placebo N = 66 n (%)	Relatives Risiko [95%-KI] p-Wert
UE	64 (94,1)	61 (92,4)	- <sup>2)</sup>
UE des NCI-CTCAE-Grades 3 oder 4	8 (11,8)	9 (13,6) <sup>3)</sup>	- <sup>4)</sup>
SUE	2 (2,9)	2 (3,0) <sup>5)</sup>	- <sup>4)</sup>
UE, das zum Abbruch der Studienmedikation oder der Studie führte	0	0	- <sup>6)</sup>

<sup>1)</sup> Alle randomisierten Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der tatsächlich erhaltenen Behandlung.

<sup>2)</sup> Patientenrelevanz von Laborparametern unklar.

<sup>3)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96 / Nachberechnungsdokument: 8 (12,1 %).

<sup>4)</sup> Aufgrund der unterschiedlichen Angaben im Nachberechnungsdokument und dem Studienbericht Woche 24 kann das vorgelegte Relative Risiko aus dem Nachberechnungsdokument nicht berücksichtigt werden.

<sup>5)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96 / Nachberechnungsdokument: 1 (1,5 %).

<sup>6)</sup> Nicht berechenbar.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; NCI: National Cancer Institute; (S)UE: (Schwerwiegendes) unerwünschte(s) Ereignis(se)

## Unerwünschte Ereignisse

Häufige UE ( $\geq 10\%$  in einem Studienarm) sind in Tabelle 23 dargestellt. Am häufigsten traten UE aus den Systemorganklassen (SOC) „Infektionen und parasitäre Erkrankungen“, „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ und „Erkrankungen des Nervensystems“ auf. Unterschiede zwischen den Studienarmen von  $\geq 5$  betroffenen Personen zeigten sich dabei im genannten SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ mit numerisch mehr Ereignissen im Placebo-Arm sowie den SOC „Erkrankungen des Nervensystems“ und „Untersuchungen“ mit jeweils numerisch mehr Ereignissen im Burosumab-Arm. Auf Ebene der *Preferred Term* (PT) traten Unterschiede von  $> 5$  betroffenen Personen bei „Arthralgie“, „Schmerzen in den Extremitäten“ und „Oropharyngeale Schmerzen“ auf. Hier entfielen jeweils numerisch mehr Ereignisse auf den Placebo-Arm.

Dargestellt wurden die Angaben aus dem Studienbericht Woche 24. Ausnahme bildet hierbei lediglich eine Abweichung im PT „Kopfschmerzen“: Im Studienbericht der Woche 24 werden hier 8 Ereignisse im Burosumab-Arm berichtet. Zusätzlich wird ein bis dato nicht in MedDRA zugeordnetes Ereignis „Kopfschmerzen“ im Burosumab-Arm angegeben. In den folgenden Studienberichten wurde dieses zuvor unkategorisierte Ereignis dem PT „Kopfschmerzen“ zugewiesen, sodass diese Angabe für Tabelle 23 übernommen wurde. Neben diesem PT sind jedoch zahlreiche weitere Diskrepanzen zwischen den Studienunterlagen festgestellt worden, für die keine Erklärung verfügbar ist. Die betreffenden Angaben sind in Tabelle 23 dargestellt.

Der pU bietet in seinem Dossier post hoc berechnete Relative Risiken und zugehörige p-Werte an. Da keine Informationen zur Teststatistik und möglichen Adjustierung vorliegen, sind diese im Anhang (siehe Tabelle 28) dargestellt.

Tabelle 23: Personen mit UE während der Behandlungsperiode mit Inzidenz  $\geq 10\%$  in einem Studienarm, Safety Set <sup>1)</sup>

MedDRA <sup>2)</sup> Systemorganklasse, Preferred Term	Burosumab N = 68 n (%)	Placebo N = 66 n (%)	Relatives Risiko [95%-KI] p-Wert
<b>Infektionen und parasitäre Erkrankungen</b>	<b>32 (47,1) <sup>3)</sup></b>	<b>30 (45,5) <sup>4)</sup></b>	- <sup>5)</sup>
Nasopharyngitis	9 (13,2) <sup>6)</sup>	6 (9,1)	- <sup>7)</sup>
Zahnabszess <sup>8)</sup>	9 (13,2)	5 (7,6) <sup>9)</sup>	- <sup>7)</sup>
<b>Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen <sup>6)</sup></b>	<b>25 (36,8) <sup>10)</sup></b>	<b>30 (45,5)</b>	- <sup>5)</sup>
Rückenschmerzen	10 (14,7)	6 (9,1)	- <sup>7)</sup>
Arthralgie	6 (8,8)	16 (24,2)	- <sup>7)</sup>
Schmerzen in den Extremitäten	5 (7,4) <sup>11)</sup>	10 (15,2)	- <sup>7)</sup>
<b>Erkrankungen des Nervensystems</b>	<b>26 (38,2)</b>	<b>16 (24,2) <sup>12)</sup></b>	- <sup>7)</sup>
Kopfschmerzen	9 (13,2) <sup>13)</sup>	5 (7,6)	- <sup>7)</sup>
Syndrom der ruhelosen Beine	8 (11,8)	4 (6,1) <sup>14)</sup>	- <sup>7)</sup>
Schwindelgefühl	7 (10,3) <sup>15)</sup>	4 (6,1) <sup>16)</sup>	- <sup>7)</sup>
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>22 (32,4) <sup>17)</sup></b>	<b>19 (28,8) <sup>18)</sup></b>	- <sup>5)</sup>
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>20 (29,4) <sup>19)</sup></b>	<b>17 (25,8) <sup>20)</sup></b>	- <sup>5)</sup>
Übelkeit	7 (10,3)	6 (9,1)	- <sup>7)</sup>

MedDRA <sup>2)</sup> Systemorganklasse, Preferred Term	Burosumab N = 68 n (%)	Placebo N = 66 n (%)	Relatives Risiko [95%-KI] p-Wert
<b>Untersuchungen</b>	<b>13 (19,1) <sup>21)</sup></b>	<b>8 (12,1) <sup>22)</sup></b>	- <sup>5)</sup>
<b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>	<b>12 (17,6)</b>	<b>14 (21,2)</b>	- <sup>7)</sup>
<b>Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>	<b>11 (16,2)</b>	<b>11 (16,7)</b>	- <sup>7)</sup>
Oropharyngeale Schmerzen	1 (1,5)	7 (10,6)	- <sup>7)</sup>
<b>Erkrankungen der Haut und des Unterhautzellgewebes</b>	<b>10 (14,7)</b>	<b>6 (9,1)</b>	- <sup>7)</sup>
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>9 (13,2)</b>	<b>6 (9,1)</b>	- <sup>7)</sup>

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der tatsächlich erhaltenen Behandlung.

<sup>2)</sup> MedDRA Version 18.1

<sup>3)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: 33 (48,5 %). Wiederum abweichende Angabe im Studienbericht Woche 96: 34 (50,0 %).

<sup>4)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 31 (47,0 %).

<sup>5)</sup> Aufgrund der unterschiedlichen Angaben im Nachberechnungsdokument und dem Studienbericht Woche 24 kann das vorgelegte Relative Risiko aus dem Nachberechnungsdokument nicht berücksichtigt werden.

<sup>6)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 11 (16,2 %).

<sup>7)</sup> Keine Angaben zur Teststatistik oder Adjustierung und im Anhang dargestellt (siehe Tabelle 31).

<sup>8)</sup> Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidität“.

<sup>9)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 6 (9,1 %).

<sup>10)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 26 (38,2 %).

<sup>11)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 6 (8,8 %).

<sup>12)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 17 (25,8 %).

<sup>13)</sup> Angabe aus Studienbericht Woche 48 / Studienbericht Woche 96 / Nachberechnungsdokument. Abweichende Angabe im Studienbericht Woche 24: 8 (11,8 %). Im Studienbericht Woche 24 ist zusätzlich ein bis dato nicht zugeordnetes Ereignis „Kopfschmerzen“ berichtet, welches in den folgenden Studienberichten dem PT „Kopfschmerzen“ zugeordnet wurde.

<sup>14)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 5 (7,6 %).

<sup>15)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 8 (11,8 %).

<sup>16)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 5 (7,6 %).

<sup>17)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: 21 (30,9 %).

<sup>18)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: 18 (27,3 %).

<sup>19)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 21 (30,9 %).

<sup>20)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 18 (27,3 %).

<sup>21)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 14 (20,6 %).

<sup>22)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 9 (13,6 %).

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; UE: Unerwünschte(s) Ereignis(se)

### UE mit NCI-CTCAE-Grad 3 oder 4

Kein UE mit Grad 3 trat bei mind. 5 % der Personen in einem Studienarm auf. Lediglich im PT „Zahnabszess“ zeigt sich eine Häufung: hier waren 3 Personen aus dem Burosumab-Arm betroffen. Kein UE wurde als Grad 4 eingestuft.

### Schwerwiegende unerwünschte Ereignisse

Insgesamt erlitten 2 Personen im Verumarm und 1 Person im Placebo-Arm ein SUE während der kontrollierten Behandlungsperiode (siehe Tabelle 16).

### Unerwünschte Ereignisse von besonderem Interesse

Im SAP waren Reaktionen an der Injektionsstelle, Hypersensibilität, Hyperphosphatämie, ektopische Mineralisierung und Syndrom der ruhelosen Beine als UE von besonderem Interesse a priori festgelegt. Im Nachberechnungsdokument werden die Ergebnisse dagegen nicht für die a priori zusammengefassten Terme, sondern auf Ebene der einzelnen PT berichtet, wodurch auf eine Darstellung im Anhang verzichtet wird.

Reaktionen an der Injektionsstelle traten bei jeweils 8 Personen in den Studienarmen während der kontrollierten Behandlungsperiode auf. Ebenfalls jeweils 4 Personen waren von Hypersensibilität betroffen. Hyperphosphatämie wurde bei 4 Personen im Verumarm und keiner Person im Placebo-Arm festgestellt. Gemäß Studienprotokoll führte dies zu Entblindung und Dosisreduktionen bei den Betroffenen (vgl. Tabelle 4). Bei 8 Personen im Burosumab-Arm und 5 Personen im Placebo-Arm wurde darüber hinaus das Syndrom der ruhelosen Beine diagnostiziert. Keine Person erlitt eine ektopische Mineralisierung während der Behandlungsperiode.

Abweichende Angaben finden sich zwischen Studienbericht Woche 24 und Studienbericht Woche 96: Laut Studienbericht Woche 96 wurde anstelle von 5 Personen bei 6 Personen im Placebo-Arm ein Syndrom der ruhelosen Beine festgestellt.

*Tabelle 24: Personen mit UE von besonderem Interesse während der Behandlungsperiode, Safety Set <sup>1)</sup>*

UE von besonderem Interesse	Burosumab N = 68 n (%)	Placebo N = 66 n (%)	Relatives Risiko [95%-KI] p-Wert
Reaktionen an der Injektionsstelle	8 (11,8)	8 (12,1)	- <sup>2)</sup>
Hypersensibilität	4 (5,9)	4 (6,1)	- <sup>2)</sup>
Hyperphosphatämie	4 (5,9)	0	- <sup>2)</sup>
Ektopische Mineralisierung	0	0	- <sup>2)</sup>
Syndrom der ruhelosen Beine	8 (11,8)	5 (7,6) <sup>3)</sup>	- <sup>2)</sup>

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der tatsächlich erhaltenen Behandlung.

<sup>2)</sup> Es sind keine Relativen Risiken für die zusammengefassten Terme (siehe Kapitel 2.3.4) verfügbar.

<sup>3)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angaben im Studienbericht Woche 96: 6 (9,1 %).

Abkürzungen: UE: Unerwünschte(s) Ereignis(se)

## 4 Diskussion der Methodik und Ergebnisse

### 4.1 Zulassungsstatus und Zulassungspopulation von Burosumab

Gemäß Fachinformation ist Burosumab indiziert zur Behandlung der XLH bei Kindern und Jugendlichen im Alter von 1 bis 17 Jahren mit röntgenologischem Nachweis einer Knochenerkrankung, und bei Erwachsenen. Die Therapie sieht bei Erwachsenen die subkutane Injektion des Antikörpers alle 4 Wochen durch eine Ärztin, einen Arzt oder eine andere geschulte medizinische Fachkraft vor. Die empfohlene Dosis wird mit 1,0 mg/kg Körpergewicht angegeben. Eine zeitgleiche Behandlung mit Phosphat und aktivem Vitamin D ist kontraindiziert [11].

Die Zulassung von Burosumab umfasste bisher die Behandlung von Kindern ab 1 Jahr und Jugendlichen in der Skelettwachstumsphase mit XLH und röntgenologischem Nachweis einer Knochenerkrankung. Am 30. September 2020 erfolgte die Zulassungserweiterung auf Erwachsene. Die vorliegende Bewertung bezieht sich gemäß Zulassungserweiterung auf erwachsene Patientinnen und Patienten mit XLH.

Die vorliegende Studie CL303 rekrutierte erwachsene XLH-Patientinnen und -Patienten im Alter von 18 bis 65 Jahren und vorhandenen Knochenschmerzen zur ersten Screeningvisite. Die Eingrenzung auf Personen mit Knochenschmerzen findet sich nicht im Anwendungsgebiet von Burosumab wieder. Der pU liefert keine Angaben dazu, bei welchem Anteil der erwachsenen XLH-Erkrankten eine Symptomatik in Form von Knochenschmerzen vorliegt. Daher ist unklar, inwiefern es sich bei der Studienpopulation um stärker betroffene Patientinnen und Patienten gegenüber der Zulassungspopulation handelt.

In der Studie CL303 nahmen keine Personen aus Deutschland teil. Die Ergebnisse der Patientinnen und Patienten aus den beteiligten Ländern erscheinen jedoch auf den deutschen Versorgungskontext übertragbar.

### 4.2 Design und Methodik der Studie CL303

Bei der Studie CL303 handelt es sich um eine randomisierte, doppelblinde, Placebo-kontrollierte, multizentrische Studie der Phase III im Parallelgruppendesign (1:1) mit 2 Behandlungsarmen. In dieser Studie wurde die Wirksamkeit und Sicherheit von Burosumab gegenüber Placebo bei Erwachsenen mit XLH untersucht. Primäres Zielkriterium der Studie war das Erreichen eines mittleren Serumphosphatspiegels oberhalb der unteren Grenze des Normalbereichs.

Die Studie gliederte sich in eine Screening- und eine 24-wöchige Behandlungsperiode. Anschließend folgte eine einarmige Behandlungsfortsetzungsphase (Woche 25 bis 48), in der alle Studienteilnehmenden unverblindet Burosumab erhielten, jedoch gegenüber ihrer vorherigen Behandlungszuteilung während der Behandlungsperiode verblindet blieben. An die Behandlungsfortsetzungsphase schlossen sich die freiwillige, offene und einarmige Extensionsperiode I für 48 Wochen (Woche 49 bis Woche 96) und für alle Studienteilnehmenden aus Nordamerika/Europa die Extensionsperiode II für bis zu 53 weitere Wochen bis zum Studienende im September 2018 an. In der Nutzenbewertung werden die Ergebnisse der 24-wöchigen, Placebo-kontrollierten Behandlungsperiode berücksichtigt.

Die Randomisierung in die Studienarme Burosumab und Placebo erfolgte stratifiziert für „Durchschnittliche Schmerzen in den letzten 7 Tagen“ (BPI-SF Frage 5  $\leq$  6,0 vs.  $>$  6,0) sowie für

Region (Nordamerika/Europa vs. Japan vs. Südkorea). Ursprünglich war eine Stratifizierung für „Schlimmste Schmerzen in den letzten 7 Tagen“ (BPI-SF Item 3  $\leq 6,0$  vs.  $> 6,0$ ) anstelle des BPI-SF Item 5 vorgesehen. Aufgrund eines Fehlers im IWRS wurde jedoch das Item 5 des BPI-SF zur Stratifizierung programmiert. Da die Rekrutierung bereits begonnen hatte, als der Fehler bemerkt wurde, wurde die stratifizierte Randomisierung mit Item 5 fortgeführt. Der pU untersuchte auf verschiedenen Wegen, inwiefern dieser Fehler die Ergebnisse der Studie beeinflusst. Unter anderem wurde überprüft, wie hoch die Übereinstimmung der Verteilungen in den jeweiligen Kategorien ist. Bei etwa 60 % der Studienteilnehmenden stimmte die tatsächliche Stratifizierungskategorie des BPI-SF Item 5 und die geplante Stratifizierungskategorie des BPI-SF Item 3 überein (z. B. indem die geplante Stratifizierungskategorie  $\leq 6,0$  ebenfalls der tatsächlichen Stratifizierungskategorie  $\leq 6,0$  entsprach). Der Pearson-Korrelationskoeffizient zwischen den Kategorien betrug 0,82 ( $p < 0,0001$ ). Dennoch zeigten sich ungleiche Verteilungen in den Kategorien des geplanten Stratifizierungsfaktors mittels Item 3 des BPI-SF in den Studienarmen: 75 % der Studienteilnehmenden aus dem Verumarm und 59 % aus dem Placebo-Arm wiesen einen Wert  $> 6$  auf. Basierend auf den Mittelwerten zeigen sich zu Baseline jedoch keine deutlichen Unterschiede zwischen den Studienarmen in diesem Merkmal. Zusammenfassend wird aufgrund dieser Aspekte daher die Einschätzung getroffen, dass die Studienarme sich nicht wesentlich hinsichtlich ihrer Schmerzausprägung zu Studienbeginn unterschieden, sodass das Verzerrungspotential aus diesem Aspekt insgesamt als niedrig eingestuft wird. Dennoch ist durch die ungleiche Verteilung in den Kategorien des geplanten Stratifizierungsfaktors des Item 3 des BPI-SF ein Einfluss auf die Ergebnisse denkbar. Der pU näherte sich dieser Frage, indem er Sensitivitätsanalysen mit Adjustierung für Item 3 anstelle des Item 5 des BPI-SF durchführte. Jedoch liegen diese nicht für alle in der Nutzenbewertung dargestellten Auswertungen vor (siehe Kapitel 4.4).

Unterschiede zwischen den Studienarmen zeigten sich auch in anderen Aspekten der Populationscharakteristika: In der XLH-spezifischen Anamnese hatten 47 der 68 Personen im Verumarm und 38 der 66 Personen im Placebo-Arm bereits Symptome einer Osteoarthritis erlebt. In der Burosumab-Gruppe waren 6 Personen mehr bereits an einer Nephrokalzinose erkrankt als in der Placebo-Gruppe. Bis zum Datenschnitt am 22. Dezember 2016 wiesen 60 Studienteilnehmende wesentliche Protokollverletzungen auf. Davon waren 10 Personen mehr im Burosumab-Studienarm als im Placebo-Studienarm betroffen. Es ist jedoch unklar, welcher Anteil der Protokollverletzungen während der kontrollierten Behandlungsperiode bis Woche 24 auftrat, da keine Auswertung für diese Studienperiode identifiziert werden konnte.

Auch die Begleitmedikation wurde in der Studie CL303 analog zu den Protokollverletzungen nur für die Datenschnitte, nicht für die kontrollierte Behandlungsperiode berichtet. Daher sind auch die Anteile an Begleitmedikation für die kontrollierte Behandlungsperiode unbekannt. Bis zum Datenschnitt am 22. Dezember 2016 wurden Differenzen  $> 10$  Prozentpunkte in den Anteilen zwischen den Studienarmen für Analgetika, Antiphlogistika und Antirheumatika, Vitamine sowie Husten- und Erkältungsmittel mit höheren Anteilen jeweils für den Verumarm beobachtet. Es ist unklar, welche Gründe für die Einnahme der Begleitmedikation in beiden Studienarmen vorlagen und ob sich diese Gründe zwischen den Studienarmen unterschieden. Schmerzmittel wurden zudem über ein Patiententagebuch, welches in der Woche vor der Studienvisite zum Einsatz kam, erfasst. Diese Erhebung wurde für die kontrollierte Behandlungsperiode ausgewertet. Zu Woche 24 hatten 45 Personen aus dem Burosumab-Arm und 40 Personen aus dem Placebo-Arm die Einnahme von Schmerzmitteln in der Woche vor Studienvisite berichtet.

Aufgrund des geringen Stichprobenumfangs erscheinen die dargestellten Unterschiede zwischen den Studienarmen jedoch nicht ungewöhnlich und können zufallsbedingt aufgetreten sein.

Hinsichtlich der Aufrechterhaltung der Verblindung waren Ausnahmeregelungen beim Überschreiten der oberen Grenze des Normalbereichs des Serumphosphatspiegels vorgesehen, in welchem Fall Dosisreduktionen mit einhergehender Entblindung vorgesehen waren. Die Angaben zur Häufigkeit dieser Entblindungen in den Studienunterlagen sind allerdings nicht konsistent und schwanken zwischen 4 bis 6 Personen im Burosumab-Arm. Im EPAR findet sich zudem die Aussage, dass lediglich das Studienpersonal, nicht die betroffene Testperson entblindet wurde. Zusätzlich wird im EPAR berichtet, dass es zu 2 versehentlichen Entblindungen jeweils im Burosumab-Arm kam. Aus den Studienunterlagen wird hingegen 1 Entblindung als Protokollverletzung berichtet. Dabei ist es auch auf Basis des EPAR abschließend unklar, ob die Entblindung auch die Testperson betraf. Insgesamt betreffen diese Aspekte jedoch nur wenige Studienteilnehmende, sodass davon ausgegangen wird, dass die Verblindung für den Großteil der Studienpopulation aufrechterhalten werden konnte.

Die Therapie mit Burosumab wurde in der Studie CL303 gemäß Fachinformation mit einer Dosis von 1,0 mg/kg alle 4 Wochen durchgeführt. Während der kontrollierten Behandlungsperiode waren nur wenige Dosisanpassungen notwendig, sodass die mittleren Dosen keine Schwankungen aufweisen.

Das Studienprotokoll der Studie CL303 wurde durch 5 Protokollamendments nach Studienbeginn modifiziert. Unklarheiten bestehen hinsichtlich der Einführung von Denosumab und Teriparatid als Ausschlusskriterium und nicht erlaubter Begleitmedikation mit Amendment 1. Der pU macht keine Angabe, ob bereits Personen in die Studie eingeschlossen waren, die eine solche Therapie eingenommen hatten, und wie mit diesen Personen umgegangen wurde. Weitere Unklarheiten existieren zur Aufnahme des Stratifizierungsfaktors „Region“ mit Amendment 3. Zum Zeitpunkt des ersten Studienberichts mit Datenschnitt vor Amendment 3 waren bereits alle Studienteilnehmenden rekrutiert. Zudem scheint die Änderung der stratifizierten Randomisierung während der laufenden Studie auch hinsichtlich der Argumentation zum Beibehalten des eigentlich fehlerhaften Stratifizierungsfaktors des Item 5 aus dem BPI-SF unwahrscheinlich. Es wird daher angenommen, dass die Aufnahme des Stratifizierungsfaktors „Region“ in Amendment 3 einer nachträglichen Dokumentation der bereits zuvor durchgeführten stratifizierten Randomisierung entspricht. Damit ergäbe sich, dass bis zum Amendment 3 die Studiendurchführung nicht nach Studienprotokoll stattfand. Dieses Vorgehen erscheint ungewöhnlich für eine klinische Studie der Phase III und erhöht insgesamt die Unsicherheit in der Einschätzung der Ergebnisse.

Der pU hat ein COA-Dossier eingereicht, welches die Auswahl der relevanten PRO-Instrumente im Anwendungsgebiet der XLH ausführlich beschreibt. Zudem sind eigene Untersuchungen zu den psychometrischen Eigenschaften der Instrumente im Erkrankungsbild dargestellt. Der ausführliche, konzeptuelle Umriss und die spezifischen Untersuchungen der Instrumente werden ausdrücklich begrüßt. Sie ermöglichen eine detaillierte Einschätzung zur Geeignetheit der Instrumente, die andernfalls nicht möglich gewesen wäre.

Die statistische Analyse wurde im SAP geplant. Der SAP wurde durch 2 Amendments nach bereits erfolgtem primären Datenschnitt modifiziert. Die Änderungen der Amendments betreffen zusätzlich eingeführte Sensitivitäts- und Subgruppenanalysen und die Aufnahme des Faktors „Region“ in die Analyse der Wirksamkeitseindpunkte. Hierbei ist auffällig, dass bereits die Analysen auf Basis des ersten Datenschnitts für „Region“ adjustiert waren, obwohl dies gemäß der damals gültigen Version

des SAP nicht vorgesehen war. Ebenfalls finden sich bereits im ersten Studienbericht die Sensitivitätsanalysen mit den Imputationsmethoden LOCF, BOCF und mBOCF, welche erst im folgenden SAP-Amendment geplant wurden. Ungewöhnlich erscheint in Anbetracht des erwarteten Stichprobenumfangs auch die Wahl von GEE-Modellen für die kontinuierlichen Auswertungen. Im EPAR ist berichtet, dass die Modellannahme der Normalverteilung für die Endpunkte des Item 3 des BPI-SF, WOMAC „Steifheit“ und WOMAC „Physische Funktion“ nicht erfüllt war. Im EPAR werden vor diesem Hintergrund ANCOVA-Auswertungen diskutiert, da eine Überprüfung der Normalverteilung der Residuen stattfand. Es liegen keine Angaben zu den Ergebnissen dieser ANCOVA-Auswertungen vor. Die Analysen selbst wurden vom pU im Dossier nicht eingereicht. Aufgrund fehlender Angaben ist zudem unklar, inwiefern die Annahme für weitere im Dossier und in der Nutzenbewertung dargestellte Endpunkte erfüllt war.

Der pU legt mit Modul 4 Subgruppenanalysen zu allen im Dossier dargestellten Endpunkten vor. Dabei wurden auch nicht a priori angedachte Subgruppenmerkmale ausgewertet. Der pU begründet die Auswahl der post hoc gewählten Subgruppenmerkmale nicht. Aufgrund von Abweichungen zu den Studienunterlagen konnten die Subgruppenanalysen zudem nur für einen geringen Teil der in der Nutzenbewertung dargestellten Endpunkte herangezogen werden, sodass auf Basis der verfügbaren Datengrundlage keine belastbaren Aussagen zu möglichen Effektmodifikationen getroffen werden können.

Das Verzerrungspotential auf Studienebene wird insgesamt mit niedrig bewertet. Maßgeblich hierfür ist das randomisiert-kontrollierte, doppelblinde Studiendesign. Es wird davon ausgegangen, dass die Studienarme trotz des Fehlers in der stratifizierten Randomisierung vergleichbar sind und dass die Verblindung für den Großteil der Studienpopulation aufrechterhalten werden konnte. Unsicherheiten hinsichtlich des Verzerrungspotentials ergeben sich auf Studienebene dennoch aus der fehlenden Angabe zur Methode der Erzeugung der Randomisierungssequenz. Die diskutierten Aspekte der statistischen Analyse und Abweichungen der Studienunterlagen werden differenziert auf Endpunktebene betrachtet, sodass eventuelle Auswirkungen auf das Verzerrungspotential in den folgenden Kapiteln verortet sind.

### **4.3 Mortalität**

Todesfälle wurden im Rahmen der Sicherheitserfassung dokumentiert. Im Verlauf der Studie CL303 ist keine Person verstorben.

### **4.4 Morbidität**

Der pU legt in seinem Dossier u. a. Ergebnisse zum Laborparameter Serumphosphat vor, welcher jedoch keinen unmittelbar patientenrelevanten Endpunkt darstellt (siehe Kapitel 2.3.2). Als klinisch bedeutsamer Parameter wurden die Ergebnisse in der Nutzenbewertung ergänzend dargestellt.

Zur Bewertung der Wirksamkeit in der Kategorie Morbidität wurden die Endpunkte Gehfähigkeit mittels 6MWT, Schmerzen, Steifheit und physische Funktion mittels WOMAC, Schmerzintensität und Schmerzbeeinträchtigung mittels BPI-SF, Fatigue mittels BFI sowie die patientenberichtete Veränderung des Krankheitszustands mittels PGI-I herangezogen.

Der pU begründet nicht nachvollziehbar, warum 2 Instrumente zur Erhebung von Schmerzen in der Studie CL303 zum Einsatz kamen. Es ist auch unklar, warum in der Auswertung dem BPI-SF gegenüber der Subskala „Schmerzen“ des WOMAC der Vorzug gegeben wurde.

**Gehfähigkeit: 6MWT**

Es zeigte sich in der primären Analyse kein statistisch signifikanter Unterschied in den Veränderungen der absoluten Gehstrecke von Baseline zu Woche 24 zwischen den Studienarmen.

Zu einem anderen Ergebnis kommt eine spätere Auswertung, in der der Baseline-Wert einer Testperson nach bereits erfolgtem Datenbankschluss nachträglich verändert wurde. Die Änderung wird damit begründet, dass es sich um die Korrektur eines fehlerhaften Baseline-Werts handelte, der auf einen Berechnungsfehler im Studienzentrum zurückzuführen sei. In der Konsequenz verändert sich das Maximum der Baseline-Werte im Burosumab-Studienarm von 860 Meter auf 643 Meter. Unklar ist, wie der fehlerhafte Wert nach bereits erfolgter Datenbereinigung für den ursprünglichen Datenschluss entdeckt wurde und worin der Berechnungsfehler konkret bestand. Da die Änderung zudem zwangsläufig nach Entblindung und unter Kenntnis der Ergebnisse aus der ersten Analyse stattfand, werden die Ergebnisse der ersten Analyse in der Nutzenbewertung berücksichtigt. In seinem Dossier präsentiert der pU zudem eine post hoc durchgeführte Sensitivitätsanalyse, in der die Person mit dem betreffenden Baseline-Wert von der Analyse ausgeschlossen wurde. Hier zeigte sich ein statistisch signifikanter Unterschied zwischen den Studienarmen zugunsten von Burosumab. Auffällig ist, dass die Veränderung eines Baseline-Werts die Signifikanz des Ergebnisses ändert, sodass die Analyse grundsätzlich wenig robust erscheint.

Keine Angaben wurden zu den Gründen für fehlende Werte in der Auswertung des 6MWT identifiziert. Da diese lediglich < 5 % der FAS-Population betrafen, wird das Verzerrungspotential aus diesem Aspekt jedoch dennoch als gering eingestuft.

Relevante Unsicherheiten ergeben sich jedoch daraus, dass nicht bekannt ist, ob die Annahme der Normalverteilung für die GEE-Analysen erfüllt war. Zudem ist keine Sensitivitätsanalyse mit Adjustierung für den geplanten Stratifizierungsfaktor (Item 3 des BPI-SF) anstelle des tatsächlichen Stratifizierungsfaktors vorhanden. In der Gesamtschau wurde das Verzerrungspotential für den Endpunkt als hoch eingestuft. Der vom pU ebenfalls vorgeschlagene Normwertvergleich (Prozentsatz der vorhergesagten Normalstrecke) wird nicht berücksichtigt, da die Repräsentativität der zugrunde liegenden Normstichproben aufgrund der willkürlichen Samplingstrategie (Convenience Sample) nicht gegeben ist.

**Schmerzen, Steifheit und Physische Funktion: WOMAC**

WOMAC wurde zur patientenberichteten Beurteilung von Schmerzen, Steifheit und physischer Funktionsfähigkeit in der Studie CL303 eingesetzt. Das Instrument wurde bis zum Ende der kontrollierten Behandlungsperiode in 2 Studienvisiten nach Baseline (Woche 12, Woche 24) eingesetzt. Mit einem Bezugszeitraum von 48 Stunden erlauben die Erhebungen daher nur Aussagen für einen geringen Anteil der Studienzzeit.

Die Subskalen von WOMAC wurden einerseits als Gruppenunterschied in den mittleren Veränderungen von Baseline zu Woche 24 sowie für die Subskalen „Steifheit“ und „Physische Funktion“ als Responderanalysen ausgewertet. In der Nutzenbewertung sind die Responderanalysen jedoch nicht dargestellt, da die Schwellenwerte als nicht ausreichend valide bewertet wurden. Problematisch erscheint zudem, dass der Einsatz von WOMAC nicht in der Studienvisite bei vorzeitigem Abbruch vorgesehen war. Da lediglich eine Person aus dem Burosumab-Arm die Studie vor Ende der Behandlungsperiode zu Woche 24 abgebrochen hatte, führt dies jedoch zu keinen weiteren Einschränkungen in der Validität der Ergebnisse.

Es zeigten sich keine statistisch signifikanten Unterschiede in den Veränderungen von Baseline zu Woche 24 in der Subskala „Schmerzen“. Dabei handelt es sich um das Ergebnis einer Post-hoc-Analyse für das Modul 4 des Dossiers, da die Subskala „Schmerzen“ in der Studie CL303 zwar erhoben wurde, a priori aber nicht zur Auswertung vorgesehen war. Der pU begründet dies mit der Erhebung von Schmerzen anhand eines weiteren Instruments, dem BPI-SF. Es bleibt offen, warum keine Erkenntnisse aus der Auswertung der Subskala „Schmerzen“ des WOMAC erwartet wurden.

Unklar ist zudem, woher Unterschiede in den Baseline-Werten für den Placebo-Arm in den statistischen Outputs aus den Studienberichten und dem Nachberechnungsdokument rühren. Für den Burosumab-Arm finden sich keine Abweichungen. Für die Nutzenbewertung ist lediglich die Post-hoc-Auswertung auf Basis des Baseline-Werts aus dem Studienbericht Woche 96 und Nachberechnungsdokument verfügbar.

In der Subskala „Steifheit“ zeigte sich im Studienbericht Woche 24 mit den Ergebnissen des ersten Datenschnitts ein statistisch signifikanter Unterschied in den Veränderungen von Baseline zu Woche 24 zugunsten von Burosumab. In den späteren Studienberichten Woche 48 und Woche 96 sowie im Dossier wird eine andere, jedoch vergleichbare Auswertung dargestellt. Für die Auswertung in den späteren Studienberichten wurde die Datenbank nach Datenbankschluss erneut geöffnet. Jedoch ist in den Studienunterlagen nicht beschrieben, welche Änderung aus welchem Grund durchgeführt wurde. Da die Änderung nach Entblindung und unter Kenntnis der Ergebnisse aus der ersten Auswertung geschah, werden die Ergebnisse der ursprünglichen Auswertung aus dem ersten Studienbericht Woche 24 für die Nutzenbewertung herangezogen. Für diese Auswertung ist jedoch kein Hedges g zur Einschätzung der klinischen Relevanz verfügbar.

Für die Subskala „Physische Funktion“ wurden keine Abweichungen in den Studienunterlagen identifiziert. Hier zeigt sich in allen Studienunterlagen der gleiche statistisch signifikante Unterschied in den Veränderungen von Baseline zu Woche 24 zugunsten von Burosumab. Gemäß Hedges g bleibt die klinische Relevanz des Effekts jedoch unklar, da das Konfidenzintervall nicht vollständig außerhalb von  $|0,2|$  liegt.

Sensitivitätsanalysen liegen lediglich für die Subskalen „Steifheit“ und „Physische Funktion“ vor. Interaktionstests zur Auswertung der Subskala „Steifheit“ aus dem ersten Studienbericht Woche 24 wurden nicht vorgelegt. Für die Subskalen „Physische Funktion“ und „Schmerzen“ zeigten sich keine Belege für eine Effektmodifikation durch die untersuchten Subgruppenmerkmale.

Keine Angaben wurden zu den Gründen für fehlende Werte in den Auswertungen der WOMAC-Subskalen identifiziert. Da der Anteil an fehlenden Werten lediglich  $< 5\%$  der FAS-Population betrifft, wird das Verzerrungspotential aus diesem Aspekt jedoch dennoch als gering eingestuft.

Das Verzerrungspotential wurde insgesamt mit unklar bewertet. Neben den bereits diskutierten Aspekten ist hierfür maßgeblich, dass laut EPAR die Annahme der Normalverteilung für die Subskalen „Physische Funktion“ und „Steifheit“ nicht erfüllt war. Alternativ wurden für den EPAR ebenfalls ANCOVA-Auswertungen vorgelegt, die nicht im Rahmen des Dossiers eingereicht wurden. Es liegen keine Informationen dazu vor, ob die Annahme der Normalverteilung für die Subskala „Schmerzen“ erfüllt war.

### **Schmerzintensität und Schmerzbeeinträchtigung: BPI-SF**

Anhand des BPI-SF wurden in der Studie CL303 Schmerzen patientenberichtet erfasst. Das Instrument wurde bis zum Ende der kontrollierten Behandlungsperiode in 2 Studienvisiten nach Baseline (Woche 12, Woche 24) eingesetzt. Wie auch bereits bei WOMAC angemerkt, wird der

fehlende Einsatz des Instruments in der Studiervisite bei vorzeitigem Abbruch grundsätzlich kritisch gesehen, führt aufgrund lediglich eines Studienabbruchs in der Behandlungsperiode jedoch zu keinen weiteren Einschränkungen in der Validität der Ergebnisse.

In der Ergebnisinterpretation des BPI-SF muss die unterschiedliche Erhebung der Domäne „Schmerzintensität“ (einschließlich des Item 3) und der Domäne „Schmerzbeeinträchtigung“ berücksichtigt werden: Die Items der Domäne „Schmerzintensität“ wurden in den 7 Tagen vor einer Studiervisite täglich durch ein papierbasiertes Patiententagebuch erfasst und zudem während der Studiervisiten erhoben. Items der Domäne „Schmerzbeeinträchtigung“ wurden lediglich während der Studiervisiten erhoben. Es wurde keine Rationale für die unterschiedlichen Erhebungsformen identifiziert. Mit einem Bezugszeitraum von 24 Stunden erlaubt das Instrument insgesamt nur für einen geringen Anteil der Studienzeit Aussagen. Dabei bildet die Domäne „Schmerzintensität“ durch die Erhebung in der Woche vor den Studiervisiten jedoch einen deutlich größeren Zeitraum ab als die Domäne „Schmerzbeeinträchtigung“.

Es zeigte sich kein statistisch signifikanter Unterschied zwischen den Studienarmen in der Responderanalyse zum Item 3 („Schlimmste Schmerzen“) für Verringerungen um  $> 15\%$  von Baseline als Schwellenwert für die geringste klinisch relevante Veränderung. Dieses Ergebnis wird durch die ergänzend dargestellte Auswertung mit Verringerung  $> 30\%$  von Baseline mit ebenfalls nicht statistisch signifikantem Gruppenunterschied unterstützt. Für diese Responderanalysen liegen lediglich post hoc berechnete Effektschätzer vor, für welche keine Angaben zur Teststatistik und möglichen Adjustierung identifiziert wurden. Sie werden daher für die Nutzenbewertung nicht herangezogen und sind lediglich im Anhang dargestellt.

Unklar ist, warum der pU für eine ausgewogene Ergebnisinterpretation nicht ebenfalls die Anteile der Personen mit Verschlechterungen ausgewertet hat.

Es zeigten sich ebenfalls keine statistisch signifikanten Unterschiede in den Veränderungen von Baseline zu Woche 24 in den Domänen „Schmerzintensität“ und „Schmerzbeeinträchtigung“. Es ist unklar, ob die Annahme der Normalverteilung für diese GEE-Auswertungen erfüllt war. Begründete Zweifel rühren aus der Aussage im EPAR, dass die Annahme der Normalverteilung in der Einzelauswertung des Item 3 „Schlimmste Schmerzen“ nicht erfüllt war.

Unsicherheiten ergeben sich für die Auswertungen der mittleren Veränderungen zudem aus inkonsistent abweichenden Angaben in den Studienunterlagen, die jedoch die Ergebnisinterpretation nicht beeinflussten. Berücksichtigt wurden die Auswertungen aus dem ersten Studienbericht Woche 24.

Keine Angaben wurden zu den Gründen für fehlende Werte in diesen Auswertungen der Domänen des BPI-SF identifiziert. Da der Anteil an fehlenden Werten lediglich  $< 5\%$  der FAS-Population betrifft, wird das Verzerrungspotential aus diesem Aspekt jedoch dennoch als gering eingestuft.

Sensitivitätsanalysen für die Responderanalysen sowie die Domänen „Schmerzintensität“ und „Schmerzbeeinträchtigung“ wurden weder geplant noch durchgeführt. Aufgrund der Unterschiede zwischen Nachberechnungsdokument und Analysen aus dem Studienbericht Woche 24 können die vorgelegten Interaktionstests aus dem Nachberechnungsdokument nicht berücksichtigt werden. Mögliche Effektmodifikationen durch die Subgruppenmerkmale können daher nicht beurteilt werden.

Unter Berücksichtigung der diskutierten Aspekte wird das Verzerrungspotential als unklar bewertet.

**Fatigue: BFI**

Fatigue wurde in der Studie CL303 mittels BFI erhoben. Auch dieses Instrument wurde bis zum Ende der kontrollierten Behandlungsperiode in 2 Studienvisiten nach Baseline (Woche 12, Woche 24) und nicht in der Studienvisite bei vorzeitigem Abbruch eingesetzt. Da es lediglich zu einem Studienabbruch in der Behandlungsperiode kam, führt der fehlende Einsatz in der Studienvisite bei vorzeitigem Abbruch jedoch zu keinen weiteren Einschränkungen in der Validität der Ergebnisse.

Wie auch der BPI-SF, ist dieses Instrument in 2 Domänen aufgebaut, die in der Studie CL303 unterschiedlich erhoben wurden: Während die Items der „Stärke der Fatigue“ in der Woche vor einer Studienvisite zusätzlich täglich über ein Patiententagebuch erfasst wurden, wurden die Items zur „Beeinträchtigung durch Fatigue“ ausschließlich während der Studienvisiten erhoben. Auswertungen wurden vorgelegt für die 3 Einzelitems der Domäne „Stärke der Fatigue“ sowie für den Gesamtwert des BFI.

Im Gegensatz zum BPI-SF scheint für das Instrument lediglich die Auswertung eines Gesamtwerts, nicht jedoch von den Einzelitems vorgesehen zu sein. Laut Entwicklungsstudie [24] erhebt der BFI ein einzelnes Konstrukt. In der Nutzenbewertung wird ausschließlich der Gesamtwert berücksichtigt. Bei der Berechnung des Gesamtwerts wurden die Erhebungen aus dem Patiententagebuch nicht berücksichtigt, sodass die Fatigue-Symptomatik der letzten 24 Stunden vor den Studienvisiten Woche 12 und Woche 24 abgebildet wird. Damit erlaubt der BFI-Gesamtwert in der Studie CL303 nur Aussagen für einen geringen Anteil der Behandlungsperiode. Es konnte keine Begründung des pU für das gewählte Vorgehen in Bezug auf die unterschiedlichen Erhebungsformen und Auswertung identifiziert werden.

Es zeigten sich keine statistisch signifikanten Unterschiede zwischen den Studienarmen in den Veränderungen von Baseline zu Woche 24 im Gesamtwert des BFI. In den Studienunterlagen wurden auch hier voneinander abweichende Angaben festgestellt, die jedoch die Signifikanz der Ergebnisse nicht beeinflussten. Berücksichtigt wurde die Auswertung aus dem ersten Studienbericht Woche 24. Es ist unklar, ob die Annahme der Normalverteilung für diese GEE-Auswertung erfüllt war.

Keine Angaben wurden zu den Gründen für fehlende Werte für jeweils eine Person pro Studienarm identifiziert. Aufgrund des geringen Anteils im Vergleich zur FAS-Population wird das Verzerrungspotential aus diesem Aspekt jedoch dennoch als gering eingestuft.

Sensitivitätsanalysen zum Gesamtwert des BFI wurden weder geplant noch durchgeführt. Aufgrund der Unterschiede zwischen Nachberechnungsdokument und Analysen aus dem Studienbericht Woche 24 können die vorgelegten Interaktionstests aus dem Nachberechnungsdokument nicht berücksichtigt werden. Mögliche Effektmodifikationen durch die Subgruppenmerkmale können daher nicht beurteilt werden.

In der Gesamtschau wird das Verzerrungspotential als unklar bewertet.

**Patientenberichtete Veränderung des Krankheitszustands: PGI-I**

Anhand der PGI-I wurde die wahrgenommene Veränderung des Krankheitszustands seit Beginn der Studie erhoben. Das Instrument wurde nicht in der Studienvisite bei vorzeitigem Abbruch eingesetzt. Da es lediglich zu einem Studienabbruch in der Behandlungsperiode kam, führt der fehlende Einsatz in der Studienvisite bei vorzeitigem Abbruch jedoch zu keinen weiteren Einschränkungen in der Validität der Ergebnisse. Zu Baseline wurde in der Studie CL303 die

gegenwärtige Krankheitsschwere über die PGI-S bewertet. Aufgrund fehlender Angaben ist es unklar, welche Relevanz diese Erhebung im weiteren Studienverlauf hatte.

Es zeigten sich keine statistisch signifikanten Unterschiede zwischen den Studienarmen in der Bewertung der patientenberichteten Veränderung zu Woche 24. In den Studienunterlagen wurden auch hier voneinander abweichende Angaben festgestellt, die jedoch die Signifikanz der Analyse nicht beeinflussten. Berücksichtigt wurde die Auswertung aus dem ersten Studienbericht Woche 24. Es ist unklar, ob die Annahme der Normalverteilung für diese GEE-Auswertung erfüllt war.

Keine Angaben wurden zu den Gründen für fehlende Werte für jeweils eine Person pro Studienarm identifiziert. Aufgrund des geringen Anteils im Vergleich zur FAS-Population wird das Verzerrungspotential aus diesem Aspekt jedoch dennoch als gering eingestuft.

Sensitivitätsanalysen wurden für die PGI-I weder geplant noch durchgeführt. Aufgrund der Unterschiede zwischen Nachberechnungsdokument und Analysen aus dem Studienbericht Woche 24 können die vorgelegten Interaktionstests aus dem Nachberechnungsdokument nicht berücksichtigt werden. Mögliche Effektmodifikationen durch die Subgruppenmerkmale können daher nicht beurteilt werden.

In der Gesamtschau wird das Verzerrungspotential als unklar bewertet.

### **Zusammenfassende Einschätzung zur Morbidität**

Es zeigten sich bei unklarem Verzerrungspotential statistisch signifikante Vorteile für Burosumab gegenüber Placebo in den Subskalen „Steifheit“ und „Physische Funktion“ des WOMAC. Die klinische Relevanz dieser Effekte ist unklar.

Es zeigten sich keine statistisch signifikanten Unterschiede zwischen den Studienarmen bei hohem Verzerrungspotential in der Gehstrecke des 6MWT sowie bei unklarem Verzerrungspotential in der Subskala „Schmerzen“ des WOMAC, den Domänen „Schmerzintensität“ und „Schmerzbeeinträchtigung“ des BPI-SF, der Responderanalyse anhand Item 3 des BPI-SF, dem Gesamtwert des BFI sowie im PGI-I.

Inwiefern sich mögliche Verzerrungen in der Gesamtschau zugunsten oder zuungunsten von Burosumab auswirken, kann nicht abgeschätzt werden. Die Ergebnissicherheit ist aufgrund der geringen Studiengröße grundsätzlich limitiert.

## **4.5 Lebensqualität**

Lebensqualität wurde in der Studie CL303 nicht erfasst.

## **4.6 Sicherheit**

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

Kritisch angemerkt wird für die Sicherheitsendpunkte die enorm hohe Anzahl an Abweichungen in den Studienunterlagen. Diese scheinen in der Gesamtschau nicht zu anderen Ergebnisinterpretationen zu führen, da sie jeweils nur einzelne Personen betreffen, sie schwächen jedoch dennoch die Vertrauenswürdigkeit der Ergebnisse. Die Abweichungen bestehen zwischen allen 3 Studienberichten sowie auch zum Nachberechnungsdokument, welches für das Modul 4 des Dossiers angefertigt wurde. In der Regel wurde der Studienbericht der Woche 24 als Grundlage für die Darstellung der Sicherheitsergebnisse in der Nutzenbewertung herangezogen.

Unsicherheiten ergeben sich zudem daraus, dass für den Studienbericht Woche 48 und Woche 96 nachträgliche Änderungen an der Datenbank nach bereits erfolgtem Datenbankschluss berichtet werden. Es konnten keine Angaben zu den konkreten Änderungen und deren Rationale identifiziert werden. Da diese Änderung nach Entblindung und unter Kenntnis der Ergebnisse aus der ersten Auswertung geschah, werden die Ergebnisse der ursprünglichen Auswertung aus dem ersten Studienbericht Woche 24 für die Nutzenbewertung herangezogen. Die festgestellten Abweichungen zwischen den Studienberichten der Woche 48 und der Woche 96 sowie zum Nachberechnungsdokument können daraus nicht erklärt werden.

Innerhalb der Behandlungsperiode bis Woche 24 trat nahezu bei allen Patientinnen und Patienten mindestens ein UE auf. Im Kontrollarm trat ein Ereignis mehr mit Schweregrad 3 als im Interventionsarm auf. Lediglich im PT „Zahnabszess“ zeigte sich hier eine Häufung: es waren 3 Personen aus dem Burosumab-Arm und keine Person aus dem Placebo-Arm betroffen. Es trat kein UE mit Schweregrad 4 auf. Im Interventionsarm und Kontrollarm wurden jeweils 2 SUE verzeichnet. Die vorgelegten, post hoc berechneten Effektschätzer können aufgrund von Diskrepanzen zwischen Studienbericht Woche 24 und Nachberechnungsdokument nicht berücksichtigt werden. Kein UE führte zum Abbruch der Studie oder Studienmedikation.

Hinsichtlich von UE jedweden Schweregrades zeigten sich Unterschiede zwischen den Studienarmen von  $\geq 5$  betroffenen Personen im SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ mit numerisch mehr Ereignissen im Placebo-Arm sowie den SOC „Erkrankungen des Nervensystems“ und „Untersuchungen“ mit jeweils numerisch mehr Ereignissen im Burosumab-Arm. Auf Ebene der PT traten Unterschiede bei „Arthralgie“, „Schmerzen in den Extremitäten“ und „Oropharyngeale Schmerzen“ mit numerisch jeweils mehr Ereignissen im Placebo-Arm auf.

Im SAP waren „Reaktionen an der Injektionsstelle“, „Hypersensibilität“, „Hyperphosphatämie“, „ektopische Mineralisierung“ und „Syndrom der ruhelosen Beine“ als UE von besonderem Interesse a priori festgelegt. „Reaktionen an der Injektionsstelle“ und „Hypersensibilität“ traten in gleicher Anzahl in den Studienarmen auf. „Hyperphosphatämie“ wurde hingegen ausschließlich bei 4 Personen im Verumarm festgestellt. Bei 3 Personen mehr im Burosumab-Arm im Vergleich zum Placebo-Arm wurde das „Syndrom der ruhelosen Beine“ diagnostiziert. Keine Person erlitt eine „ektopische Mineralisierung“ während der Behandlungsperiode.

In den Sicherheitserhebungen wurden UE des SOC „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ sowie der PT „Zahnabszess“ und „Karies“ berichtet. Ereignisse dieser Kategorien sind jedoch auch typische Symptome der hier betrachteten Grunderkrankung, welche der Endpunktkategorie „Morbidity“ zuzuordnen wären. Es ist allerdings nicht möglich, nachträglich einzelne Symptome wie Schmerzen in den Extremitäten als Ereignis der Morbidity oder Sicherheit zu bestimmen.

Der EPAR [10] kommt zu dem Schluss, dass das Sicherheitsprofil von Burosumab für Erwachsene dem bekannten Sicherheitsprofil in der pädiatrischen Population entspricht. Es wurden keine spezifischen Auflagen zur Sicherheit formuliert.

### **Zusammenfassende Einschätzung zur Sicherheit**

Das Verzerrungspotential wurde in der Gesamtschau mit niedrig eingestuft. Insgesamt wurden nur geringe Unterschiede im Auftreten unerwünschter Ereignisse zwischen den Studienarmen festgestellt, die sowohl numerisch zugunsten als auch zuungunsten von Burosumab auftraten. Es wurden nur wenige schwerwiegende unerwünschte Ereignisse beobachtet. Häufungen traten lediglich im PT „Zahnabszess“ sowie im UE von besonderem Interesse „Hyperphosphatämie“ jeweils zuungunsten von Burosumab auf. Es ist unklar, inwiefern einige der berichteten unerwünschten Ereignisse auch Symptome der Grunderkrankung widerspiegeln.

Eine abschließende Bewertung des Sicherheitsprofils von Burosumab im Vergleich zu Placebo ist aufgrund der geringen Stichprobe besonders in Hinblick auf schwere und schwerwiegende Ereignisse nicht möglich.

## **5 Anforderung an eine qualitätsgesicherte Anwendung**

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

## 6 Zusammenfassung der Nutzenbewertung

Burosumab ist zugelassen zur Behandlung der XLH bei Kindern und Jugendlichen im Alter von 1 bis 17 Jahren mit röntgenologischem Nachweis einer Knochenerkrankung, und bei Erwachsenen [11]. Die Nutzenbewertung von Burosumab basiert auf der Zulassungsbegründenden Studie CL303 und umfasst gemäß Zulassungserweiterung erwachsene Patientinnen und Patienten mit XLH. Es handelt sich bei der Studie CL303 um eine randomisierte, Placebo-kontrollierte Studie der Phase III.

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

Tabelle 25: Zusammenfassende Darstellung der Ergebnisse der Studie UX023-CL303 für die Behandlungsperiode

Endpunkt	Ergebnis Burosumab (N = 68) vs. Placebo (N = 66)	Effekt
<b>Mortalität</b>		
Todesfälle	Keine Ereignisse aufgetreten.	n. b.
<b>Morbidität</b>		
Veränderung der 6MWT-Distanz zwischen Baseline und Woche 24 (Meter) <sup>1)</sup> LS Mean-Differenz, [95%-KI]; p-Wert <sup>2)</sup>	11,6 [-8,9; 32,2]; 0,2671	↔
Veränderung der Symptome mittels WOMAC zwischen Baseline und Woche 24 <sup>1)</sup> LS Mean-Differenz, [95%-KI]; p-Wert <sup>2)</sup>	<u>Schmerzen</u> <sup>3)</sup> : -4,4 [-9,3; 0,6]; 0,0848 <u>Steifheit</u> <sup>4)</sup> : -8,1 [-14,5; -1,8]; 0,0122 <u>Physische Funktion</u> <sup>6)</sup> : -4,9 [-9,8; -0,1]; 0,0478	↔  ↑ <sup>5)</sup> ↑ <sup>7)</sup>
Verringerung > 15 % von Baseline im Item 3 („Schlimmste Schmerzen“) <sup>10)</sup> des BPI-SF zu Woche 24 <sup>1)</sup> Effektschätzer [95%-KI]; p-Wert <sup>11)</sup>	n. v. [n. v.]; 0,3564	↔
Veränderung der Schmerzen mittels BPI-SF zwischen Baseline und Woche 24 <sup>1)</sup> LS Mean-Differenz, [95%-KI]; p-Wert <sup>2)</sup>	<u>Schmerzintensität</u> <sup>8)</sup> : -0,4 [-0,9; 0,1]; 0,0844 <u>Schmerzbeeinträchtigung</u> <sup>9)</sup> : -0,2 [-0,7; 0,4]; 0,5476	↔ ↔
Veränderung der Fatigue mittels BFI <sup>12)</sup> zwischen Baseline und Woche 24 <sup>1)</sup> LS Mean-Differenz, [95%-KI]; p-Wert <sup>2)</sup>	0,1 [-0,5; 0,6]; 0,7912	↔
Patientenberichtete Veränderung mittels PGI-I <sup>13)</sup> zu Woche 24 <sup>1)</sup>		

Endpunkt	Ergebnis Burosumab (N = 68) vs. Placebo (N = 66)	Effekt
LS Mean-Differenz, [95%-KI]; p-Wert <sup>14)</sup>		↔
<b>Sicherheit</b>		
<i>UE des CTCAE-Grades 3/4</i> Burosumab vs. Kontrollarm, n (%)	8 (11,8) vs. 9 (13,6) <sup>15)</sup>	n. b.
<i>Schwerwiegende UE</i> Burosumab vs. Kontrollarm, n (%)	2 (2,9) vs. 2 (3,0) <sup>16)</sup>	n. b.
<i>UE, das zum Abbruch der Studie/ Studienmedikation führte</i> Burosumab vs. Kontrollarm, n (%)	Keine Ereignisse aufgetreten.	n. b.
<b>UE von besonderem Interesse</b>		
<i>Reaktionen an der Injektionsstelle</i> Burosumab vs. Kontrollarm, n (%)	8 (11,8) vs. 8 (12,1)	n. b.
<i>Hypersensibilität</i> Burosumab vs. Kontrollarm, n (%)	4 (5,9) vs. 4 (6,1)	n. b.
<i>Hyperphosphatämie</i> Burosumab vs. Kontrollarm, n (%)	4 (5,9) vs. 0	n. b.
<i>Ektopische Mineralisierung</i> Burosumab vs. Kontrollarm, n (%)	Keine Ereignisse aufgetreten.	n. b.
<i>Syndrom der ruhelosen Beine</i> Burosumab vs. Kontrollarm, n (%)	8 (11,8) vs. 5 (7,6) <sup>17)</sup>	n. b.

<sup>1)</sup> Auswertung im Full Analysis Set (Burosumab: N = 68, Kontrolle: N = 66), welches dem ITT-Prinzip entspricht.

<sup>2)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen und dem jeweiligen Baseline-Wert als Kovariate.

<sup>3)</sup> Höhere Werte korrespondieren mit höheren Schmerzen.

<sup>4)</sup> Höhere Werte korrespondieren mit stärkerer Steifheit.

<sup>5)</sup> Kein Hedges g zur Bewertung der klinischen Relevanz verfügbar.

<sup>6)</sup> Höhere Werte korrespondieren mit eingeschränkterer körperlicher Funktion.

<sup>7)</sup> Es lässt sich auf Basis von Hedges g mit -0,4 (95%-KI: -0,7; 0,0) nicht ableiten, dass ein relevanter Effekt vorliegt.

<sup>8)</sup> Höhere Werte korrespondieren mit höherer Schmerzbeeinträchtigung.

<sup>9)</sup> Höhere Werte korrespondieren mit höherer Schmerzintensität.

<sup>10)</sup> Höhere Werte korrespondieren mit höheren Schmerzen.

<sup>11)</sup> A priori festgelegter Cochran-Mantel-Haenszel-Test adjustiert für den tatsächlichen Stratifizierungsfaktor des BPI-SF 5 („Durchschnittliche Schmerzen“) und Region. Die Berechnung eines zugehörigen Effektschätzers war a priori nicht geplant.

<sup>12)</sup> Höhere Werte korrespondieren mit stärkerer Fatigue.

<sup>13)</sup> Höhere Werte korrespondieren mit Verschlechterungen.

<sup>14)</sup> GEE-Modell mit Visite, Studienarm, den Stratifizierungsfaktoren BPI-SF 5 („Durchschnittliche Schmerzen“) und Region, Interaktion Visite x Studienarm als unabhängige Variablen.

<sup>15)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96 / Nachberechnungsdokument: 8 (12,1 %).

<sup>16)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96 / Nachberechnungsdokument: 1 (1,5 %).

<sup>17)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 6 (9,1 %).

Abkürzungen: 6MWT: 6-Minuten-Gehtest; BFI: Brief Fatigue Inventory; BPI-SF: Brief Pain Inventory - Short Form; CTCAE: Common Terminology Criteria for Adverse Events; GEE: Generalized estimation equation; ITT: Intention-to-treat; KI: Konfidenzintervall; LS: Least Squares; n. b.: nicht bewertbar; n. v.: nicht verfügbar; PGI-I Patient Global Impression of Improvement; SE: Standardfehler; UE: Unerwünschte(s) Ereignis(se); WOMAC: Western Ontario and McMaster Universities Osteoarthritis

## Referenzen

1. **ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories.** ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002;166(1):111-117.
2. **Bellamy N.** WOMAC Osteoarthritis Index User Guide X. Brisbane (AUS): Bellamy; 2012.
3. **Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML.** Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology (Oxford)* 1999;38(9):870-877.
4. **Cleeland CS.** The Brief Pain Inventory User Guide [online]. Houston (USA): University of Texas M. D. Anderson Cancer Center; 2009. [Zugriff: 16.11.2020]. URL: [https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI\\_UserGuide.pdf](https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf).
5. **Cleeland CS, Ryan KM.** Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singap* 1994;23(2):129-138.
6. **Daut RL, Cleeland CS, Flanery RC.** Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. *Pain* 1983;17(2):197-210.
7. **Davies GM, Watson DJ, Bellamy N.** Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res* 1999;12(3):172-179.
8. **Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al.** Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9(2):105-121.
9. **Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N.** Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000;27(11):2635-2641.
10. **European Medicines Agency (EMA).** Crysvita (Burosumab): European public assessment report EMEA/H/C/004275/II/0010/G [online]. Amsterdam (NED): EMA; 2020. [Zugriff: 10.11.2020]. URL: [https://www.ema.europa.eu/en/documents/variation-report/crysvita-h-c-4275-ii-010-g-epar-assessment-report-variation\\_en.pdf](https://www.ema.europa.eu/en/documents/variation-report/crysvita-h-c-4275-ii-010-g-epar-assessment-report-variation_en.pdf).
11. **European Medicines Agency (EMA).** Crysvita (Burosumab): Zusammenfassung der Merkmale des Arzneimittels (EPAR Product information) [online]. Amsterdam (NED): EMA; 2020. [Zugriff: 10.11.2020]. URL: [https://www.ema.europa.eu/en/documents/product-information/crysvita-epar-product-information\\_de.pdf](https://www.ema.europa.eu/en/documents/product-information/crysvita-epar-product-information_de.pdf).
12. **Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM.** Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001;94(2):149-158.
13. **Gemeinsamer Bundesausschuss (G-BA).** Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII –

- Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Burosumab (Neubewertung nach Fristablauf: Hypophosphatämie), vom 2. April 2020 [online]. Berlin (GER): G-BA; 2020. [Zugriff: 10.11.2020]. URL: [https://www.g-ba.de/downloads/39-261-4239/2020-04-02\\_AM-RL-XII\\_Burosumab\\_D-492\\_BAnz.pdf](https://www.g-ba.de/downloads/39-261-4239/2020-04-02_AM-RL-XII_Burosumab_D-492_BAnz.pdf).
14. **Gibbons WJ, Fruchter N, Sloan S, Levy RD.** Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years. *J Cardiopulm Rehabil* 2001;21(2):87-93.
  15. **Guyatt GH, Norman GR, Juniper EF, Griffith LE.** A critical look at transition ratings. *J Clin Epidemiol* 2002;55(9):900-908.
  16. **Hanley MA, Jensen MP, Ehde DM, Robinson LR, Cardenas DD, Turner JA, et al.** Clinically significant change in pain intensity ratings in persons with spinal cord injury or amputation. *Clin J Pain* 2006;22(1):25-31.
  17. **Insogna KL, Briot K, Imel EA, Kamenický P, Ruppe MD, Portale AA, et al.** A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Trial Evaluating the Efficacy of Burosumab, an Anti-FGF23 Antibody, in Adults With X-Linked Hypophosphatemia: Week 24 Primary Analysis. *J Bone Miner Res* 2018;33(8):1383-1393.
  18. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).** Allgemeine Methoden, Version 6.0 [online]. Köln (GER): IQWiG; 05.11.2020. [Zugriff: 24.11.2020]. URL: [https://www.iqwig.de/download/Allgemeine-Methoden\\_Version-6-0.pdf](https://www.iqwig.de/download/Allgemeine-Methoden_Version-6-0.pdf).
  19. **Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.** Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-766.e761.
  20. **Kyowa Kirin.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Burosumab (Crysvita), X-chromosomale Hypophosphatämie: Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 26.10.2020.
  21. **Kyowa Kirin, Ultragenyx Pharmaceutical.** A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study with Open-Label Extension to Assess the Efficacy and Safety of KRN23 in Adults with X-linked Hypophosphatemia (XLH), (UX023-CL303, Week 96/Final Analysis): clinical study report [unveröffentlicht]. 2020.
  22. **Lin CC, Chang AP, Chen ML, Cleeland CS, Mendoza TR, Wang XS.** Validation of the Taiwanese version of the Brief Fatigue Inventory. *J Pain Symptom Manage* 2006;32(1):52-59.
  23. **McConnell S, Kolopack P, Davis AM.** The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum* 2001;45(5):453-461.
  24. **Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al.** The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. *Cancer* 1999;85(5):1186-1196.

25. **Nixon A, Doll H, Kerr C, Burge R, Naegeli AN.** Interpreting change from patient reported outcome (PRO) endpoints: patient global ratings of concept versus patient global ratings of change, a case study among osteoporosis patients. *Health Qual Life Outcomes* 2016;14:25.
26. **Norman GR, Stratford P, Regehr G.** Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50(8):869-879.
27. **Okuyama T, Wang XS, Akechi T, Mendoza TR, Hosaka T, Cleeland CS, et al.** Validation study of the Japanese version of the brief fatigue inventory. *J Pain Symptom Manage* 2003;25(2):106-117.
28. **Paramita N, Nusdwinuringtyas N, Nuhonni SA, Atmakusuma TD, Ismail RI, Mendoza TR, et al.** Validity and Reliability of the Indonesian Version of the Brief Fatigue Inventory in Cancer Patients. *J Pain Symptom Manage* 2016;52(5):744-751.
29. **Portale AA, Carpenter TO, Brandi ML, Briot K, Cheong HI, Cohen-Solal M, et al.** Continued Beneficial Effects of Burosumab in Adults with X-Linked Hypophosphatemia: Results from a 24-Week Treatment Continuation Period After a 24-Week Double-Blind Placebo-Controlled Period. *Calcif Tissue Int* 2019;105(3):271-284.
30. **Radbruch L, Sabatowski R, Elsner F, Everts J, Mendoza T, Cleeland C.** Validation of the German version of the brief fatigue inventory. *J Pain Symptom Manage* 2003;25(5):449-458.
31. **Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W.** Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain* 2004;8(4):283-291.
32. **Stratford PW, Kennedy DM.** Does parallel item content on WOMAC's pain and function subscales limit its ability to detect change in functional status? *BMC Musculoskelet Disord* 2004;5:17.
33. **Ultragenyx Pharmaceutical.** Patient-reported outcome measures clinical outcome assessment dossier, v12.0 [unveröffentlicht]. 2018.
34. **Ultragenyx Pharmaceutical.** A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study with Open-Label Extension to Assess the Efficacy and Safety of KRN23 in Adults with X-linked Hypophosphatemia (XLH), (UX023-CL303, Week 24 Analysis): clinical study report [unveröffentlicht]. 2017.
35. **Ultragenyx Pharmaceutical.** A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study with Open-Label Extension to Assess the Efficacy and Safety of KRN23 in Adults with X-linked Hypophosphatemia (XLH), (UX023-CL303, Week 48 Analysis): clinical study report [unveröffentlicht]. 2018.
36. **Wang XS, Hao XS, Wang Y, Guo H, Jiang YQ, Mendoza TR, et al.** Validation study of the Chinese version of the Brief Fatigue Inventory (BFI-C). *J Pain Symptom Manage* 2004;27(4):322-332.

37. **Wolfe F.** Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. *J Rheumatol* 2004;31(10):1896-1902.
38. **Yun YH, Wang XS, Lee JS, Roh JW, Lee CG, Lee WS, et al.** Validation study of the korean version of the brief fatigue inventory. *J Pain Symptom Manage* 2005;29(2):165-172.

## Anhang

**Tabelle 26: Personen mit mittlerem Serumphosphatspiegel  $\geq 2,5$  mg/dL in der Mitte des Dosiszyklus während der Behandlungsperiode inkl. Relativem Risiko, Full Analysis Set <sup>1)</sup>**

CL303	Burosumab N = 68	Placebo N = 66
Personen mit mittlerem Serumphosphatspiegel $\geq 2,5$ mg/dL in der Mitte des Dosiszyklus n (%)	63 (92,6)	5 (7,6)
Effektschätzer [95%-KI]; p-Wert <sup>2)</sup> Relatives Risiko [95%-KI]; p-Wert <sup>3)</sup>	n. v. [n. v.]; < 0,0001 12,2 [5,3; 28,5]; < 0,0001	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> A priori festgelegter Cochran-Mantel-Haenszel-Test adjustiert für den tatsächlichen Stratifizierungsfaktor des BPI-SF 5 („Durchschnittliche Schmerzen“) und Region. Die Berechnung eines zugehörigen Effektschätzers war a priori nicht geplant.

<sup>3)</sup> Post hoc berechnet ohne weitere Angaben zur Teststatistik oder Adjustierung.

Abkürzungen: dL: Deziliter; KI: Konfidenzintervall; mg: Milligramm; n. v.: nicht verfügbar

**Tabelle 27: Personen mit Response im BPI-SF Item 3 zu Woche 24 inkl. Relativen Risiken, Full Analysis Set <sup>1)</sup>**

CL303	Burosumab N = 68	Placebo N = 66
Personen mit Verringerung > 15 % von Baseline zu Woche 24 im BPI-SF Item 3 <sup>2)</sup> n (%)	29 (42,6)	23 (34,8)
Effektschätzer [95%-KI]; p-Wert <sup>3)</sup> Relatives Risiko [95%-KI]; p-Wert <sup>4)</sup>	n. v. [n. v.]; 0,3564 1,2 [0,8; 1,9]; 0,3800	
Personen mit Verringerung > 30 % von Baseline zu Woche 24 im BPI-SF Item 3 <sup>2) 5)</sup> n (%)	17 (25,0)	12 (18,2)
Effektschätzer [95%-KI]; p-Wert <sup>3)</sup> Relatives Risiko [95%-KI]; p-Wert <sup>4)</sup>	n. v. [n. v.]; 0,2858 1,4 [0,7; 2,7]; 0,4036	

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

<sup>2)</sup> Höhere Werte korrespondieren mit höheren Schmerzen.

<sup>3)</sup> A priori festgelegter Cochran-Mantel-Haenszel-Test adjustiert für den tatsächlichen Stratifizierungsfaktor des BPI-SF 5 („Durchschnittliche Schmerzen“) und Region. Die Berechnung eines zugehörigen Effektschätzers war a priori nicht geplant.

<sup>4)</sup> Post hoc berechnet ohne weitere Angaben zur Teststatistik oder Adjustierung.

<sup>5)</sup> Ergänzend dargestellt.

Abkürzungen: BPI-SF: Brief Pain Inventory - Short Form; KI: Konfidenzintervall; n. v.: nicht verfügbar

**Tabelle 28: Personen mit UE während der Behandlungsperiode mit Inzidenz  $\geq 10$  % in einem Studienarm inkl. Relativen Risiken, Safety Set <sup>1)</sup>**

MedDRA <sup>2)</sup> Systemorganklasse, Preferred Term	Burosumab N = 68 n (%)	Placebo N = 66 n (%)	Relatives Risiko [95%-KI] p-Wert
<b>Infektionen und parasitäre Erkrankungen</b>	<b>32 (47,1) <sup>3)</sup></b>	<b>30 (45,5) <sup>4)</sup></b>	- <sup>5)</sup>
Nasopharyngitis	9 (13,2) <sup>6)</sup>	6 (9,1)	1,5 [0,5; 3,9]; 0,5858 <sup>7)</sup>
Zahnabszess <sup>8)</sup>	9 (13,2)	5 (7,6) <sup>9)</sup>	1,7 [0,6; 4,9]; 0,3986 <sup>7)</sup>
<b>Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen <sup>8)</sup></b>	<b>25 (36,8) <sup>10)</sup></b>	<b>30 (45,5)</b>	- <sup>5)</sup>
Rückenschmerzen	10 (14,7)	6 (9,1)	1,6 [0,6; 4,2]; 0,4260 <sup>7)</sup>

MedDRA <sup>2)</sup> Systemorganklasse, Preferred Term	Burosumab N = 68 n (%)	Placebo N = 66 n (%)	Relatives Risiko [95%-KI] p-Wert
Arthralgie	6 (8,8)	16 (24,2)	0,4 [0,2; 0,9]; 0,0198 <sup>7)</sup>
Schmerzen in den Extremitäten	5 (7,4) <sup>11)</sup>	10 (15,2)	0,5 [0,2; 1,3]; 0,1787 <sup>7)</sup>
<b>Erkrankungen des Nervensystems</b>	<b>26 (38,2)</b>	<b>16 (24,2) <sup>12)</sup></b>	<b>1,6 [0,9; 2,7]; 0,0954 <sup>7)</sup></b>
Kopfschmerzen	9 (13,2) <sup>13)</sup>	5 (7,6)	1,7 [0,6; 4,9]; 0,3986 <sup>7)</sup>
Syndrom der ruhelosen Beine	8 (11,8)	4 (6,1) <sup>14)</sup>	1,9 [0,6; 6,1]; 0,3657 <sup>7)</sup>
Schwindelgefühl	7 (10,3) <sup>15)</sup>	4 (6,1) <sup>16)</sup>	1,7 [0,5; 5,5]; 0,5316 <sup>7)</sup>
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>22 (32,4) <sup>17)</sup></b>	<b>19 (28,8) <sup>18)</sup></b>	- <sup>5)</sup>
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>20 (29,4) <sup>19)</sup></b>	<b>17 (25,8) <sup>20)</sup></b>	- <sup>5)</sup>
Übelkeit	7 (10,3)	6 (9,1)	1,1 [0,4; 3,2]; 1,0000 <sup>7)</sup>
<b>Untersuchungen</b>	<b>13 (19,1) <sup>21)</sup></b>	<b>8 (12,1) <sup>22)</sup></b>	- <sup>5)</sup>
<b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>	<b>12 (17,6)</b>	<b>14 (21,2)</b>	<b>0,8 [0,4; 1,7]; 0,6654 <sup>7)</sup></b>
<b>Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>	<b>11 (16,2)</b>	<b>11 (16,7)</b>	<b>1,0 [0,5; 2,1]; 1,0000 <sup>7)</sup></b>
Oropharyngeale Schmerzen	1 (1,5)	7 (10,6)	0,1 [0,0; 1,1]; 0,0317 <sup>7)</sup>
<b>Erkrankungen der Haut und des Unterhautzellgewebes</b>	<b>10 (14,7)</b>	<b>6 (9,1)</b>	<b>1,6 [0,5; 4,9]; 0,3977 <sup>7)</sup></b>
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>9 (13,2)</b>	<b>6 (9,1)</b>	<b>1,5 [0,5; 3,9]; 0,5858 <sup>7)</sup></b>

<sup>1)</sup> Alle Patientinnen und Patienten mit  $\geq 1$  Dosis der Studienmedikation. Die Auswertung erfolgte entsprechend der tatsächlich erhaltenen Behandlung.

<sup>2)</sup> MedDRA Version 18.1

<sup>3)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: 33 (48,5 %). Wiederum abweichende Angabe im Studienbericht Woche 96: 34 (50,0 %).

<sup>4)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 31 (47,0 %).

<sup>5)</sup> Aufgrund der unterschiedlichen Angaben im Nachberechnungsdokument und dem Studienbericht Woche 24 können die vorgelegten Relativen Risiken aus dem Nachberechnungsdokument nicht berücksichtigt werden.

<sup>6)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 11 (16,2 %).

<sup>7)</sup> Post hoc berechnet ohne weitere Angaben zur Teststatistik oder Adjustierung.

<sup>8)</sup> Enthält möglicherweise Ereignisse aus der Endpunktkategorie „Morbidität“.

<sup>9)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 6 (9,1 %).

<sup>10)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 26 (38,2 %).

<sup>11)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 6 (8,8 %).

<sup>12)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 17 (25,8 %).

<sup>13)</sup> Angabe aus Studienbericht Woche 48 / Studienbericht Woche 96 / Nachberechnungsdokument. Abweichende Angabe im Studienbericht Woche 24: 8 (11,8 %). Im Studienbericht Woche 24 ist zusätzlich ein bis dato nicht zugeordnetes Ereignis „Kopfschmerzen“ berichtet, welches in den folgenden Studienberichten dem PT „Kopfschmerzen“ zugeordnet wurde.

<sup>14)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 5 (7,6 %).

<sup>15)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 8 (11,8 %).

<sup>16)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 5 (7,6 %).

<sup>17)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: 21 (30,9 %).

<sup>18)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48 und im Nachberechnungsdokument: 18 (27,3 %).

<sup>19)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 96: 21 (30,9 %).

<sup>20)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 18 (27,3 %).

<sup>21)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 14 (20,6 %).

<sup>22)</sup> Angabe aus Studienbericht Woche 24. Abweichende Angabe im Studienbericht Woche 48, Studienbericht Woche 96 und im Nachberechnungsdokument: 9 (13,6 %).

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; KI: Konfidenzinterfall; UE: Unerwünschte(s) Ereignis(se)