

**Dossier zur Nutzenbewertung  
gemäß § 35a SGB V**

*Burosumab (Crysvita<sup>®</sup>)*

Kyowa Kirin GmbH

**Modul 4 B**

*Erwachsene ( $\geq 18$  Jahre) mit X-chromosomaler  
Hypophosphatämie*

Medizinischer Nutzen und  
medizinischer Zusatznutzen,  
Patientengruppen mit therapeutisch  
bedeutsamem Zusatznutzen

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>4</b>
<b>Abbildungsverzeichnis</b> .....	<b>9</b>
<b>Abkürzungsverzeichnis</b> .....	<b>10</b>
<b>4 Modul 4 – allgemeine Informationen</b> .....	<b>13</b>
4.1 Zusammenfassung der Inhalte von Modul 4.....	14
4.2 Methodik .....	27
4.2.1 Fragestellung .....	27
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung .....	30
4.2.3 Informationsbeschaffung .....	34
4.2.3.1 Studien des pharmazeutischen Unternehmers .....	34
4.2.3.2 Bibliografische Literaturrecherche .....	34
4.2.3.3 Suche in Studienregistern/ Studienergebnisdatenbanken .....	35
4.2.3.4 Suche auf der Internetseite des G-BA.....	37
4.2.3.5 Selektion relevanter Studien .....	38
4.2.4 Bewertung der Aussagekraft der Nachweise.....	38
4.2.5 Informationssynthese und -analyse .....	40
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien .....	40
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien.....	41
4.2.5.3 Meta-Analysen.....	61
4.2.5.4 Sensitivitätsanalysen.....	62
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren.....	66
4.2.5.6 Indirekte Vergleiche .....	70
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen .....	73
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	73
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel .....	73
4.3.1.1.1 Studien des pharmazeutischen Unternehmers.....	73
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche.....	75
4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken.....	77
4.3.1.1.4 Studien aus der Suche auf der Internetseite des G-BA .....	78
4.3.1.1.5 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel.....	78
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel.....	79
4.3.1.2.1 Studiendesign und Studienpopulationen .....	79
4.3.1.2.2 Verzerrungspotenzial auf Studienebene.....	95
4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien.....	98
4.3.1.3.1 Ergebnisse patientenrelevanter Endpunkte .....	99
4.3.1.3.1.1 Serumphosphat – RCT.....	102

4.3.1.3.1.2	Gehfähigkeit – RCT.....	110
4.3.1.3.1.3	Steifheit und physische Funktion – RCT.....	120
4.3.1.3.1.4	Schmerzerhebung – RCT.....	135
4.3.1.3.1.5	Fatigue – RCT .....	147
4.3.1.3.1.6	Erfassung von Therapieeffekten – RCT .....	152
4.3.1.3.1.7	Sicherheit – RCT .....	157
4.3.1.3.2	Subgruppenanalysen – RCT.....	167
4.3.1.3.2.1	Ergebnisse für präspezifizierte Subgruppenanalysen .....	176
4.3.1.3.2.2	Ergebnisse für <i>post-hoc</i> durchgeführte Subgruppenanalysen .....	183
4.3.1.4	Liste der eingeschlossenen Studien - RCT .....	189
4.3.2	Weitere Unterlagen.....	189
4.3.2.1	Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien .....	189
4.3.2.1.1	Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche .....	190
4.3.2.1.2	Charakteristika der Studien für indirekte Vergleiche.....	190
4.3.2.1.3	Ergebnisse aus indirekten Vergleichen .....	190
4.3.2.1.3.1	<Endpunkt xxx> – indirekte Vergleiche aus RCT .....	191
4.3.2.1.3.2	Subgruppenanalysen – indirekte Vergleiche aus RCT .....	193
4.3.2.1.4	Liste der eingeschlossenen Studien – indirekte Vergleiche aus RCT... ..	193
4.3.2.2	Nicht randomisierte vergleichende Studien .....	193
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien .....	193
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	194
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien .....	195
4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien.....	195
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien .....	196
4.3.2.2.4	Liste der eingeschlossenen Studien – nicht randomisierte vergleichende Studien .....	196
4.3.2.3	Weitere Untersuchungen.....	197
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen .....	197
4.3.2.3.2	Charakteristika der weiteren Untersuchungen .....	197
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen .....	198
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen .....	198
4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen .....	198
4.3.2.3.4	Liste der eingeschlossenen Studien – weitere Untersuchungen.....	199
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	200
4.4.1	Beurteilung der Aussagekraft der Nachweise .....	200
4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	201
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht .....	214
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte .....	215
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	215
4.5.2	Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen.....	215
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen .....	215

4.5.4	Verwendung von Surrogatendpunkten .....	215
4.6	Referenzliste.....	217
<b>Anhang 4-A</b>	<b>: Suchstrategien – bibliografische Literaturrecherche .....</b>	<b>223</b>
<b>Anhang 4-B</b>	<b>: Suchstrategien – Suche in Studienregistern/ Studienergebnisdatenbanken.....</b>	<b>227</b>
<b>Anhang 4-C</b>	<b>: Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche).....</b>	<b>230</b>
<b>Anhang 4-D</b>	<b>: Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern/ Studienergebnisdatenbanken).....</b>	<b>231</b>
<b>Anhang 4-E</b>	<b>: Methodik der eingeschlossenen Studien – RCT .....</b>	<b>238</b>
<b>Anhang 4-F</b>	<b>: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten .....</b>	<b>274</b>
<b>Anhang 4-G</b>	<b>: Ergänzende Analysen .....</b>	<b>296</b>

**Tabellenverzeichnis**

	<b>Seite</b>
Tabelle 4-1: Einschlusskriterien der Studien für die Nutzenbewertung von Burosumab .....	17
Tabelle 4-2: Ausschlusskriterien der Studien für die Nutzenbewertung von Burosumab .....	18
Tabelle 4-3: Zusammenfassung der Ergebnisse der Kategorie <i>Morbidität</i> zum medizinischen Zusatznutzen von Burosumab gegenüber der zweckmäßigen Vergleichstherapie (zVT) .....	19
Tabelle 4-4: Zusammenfassung der Ergebnisse der Kategorie <i>Sicherheit</i> zum medizinischen Zusatznutzen von Burosumab gegenüber der zVT .....	24
Tabelle 4-5: Einschlusskriterien für RCT mit Burosumab als Intervention.....	31
Tabelle 4-6: Ausschlusskriterien für RCT mit Burosumab als Intervention.....	32
Tabelle 4-7: Übersicht über die patientenrelevanten Endpunkte in der Studie UX023- CL303.....	42
Tabelle 4-8: Übersicht der standardisierten Effektstärke (Cohen`s d): die Unterteilung der verschiedenen Gruppen basiert auf dem <i>PGI-I</i> (UX023-CL303) .....	51
Tabelle 4-9: Übersicht über die standardisierte Effektstärke (Cohen`s d): die Unterteilung der verschiedenen Gruppen basiert auf dem <i>6MWT</i> (UX023-CL303) .....	52
Tabelle 4-10: Nummerierung der einzelnen Items des <i>BPI</i> -Fragebogens für die Stratifizierung in der Studie UX023-CL303 .....	63
Tabelle 4-11: Kriterien für die Berechnung der Interaktionstests in Anlehnung an das Methodenpapier 6.0 des IQWiG .....	68
Tabelle 4-12: Präspezifizierte Subgruppen .....	68
Tabelle 4-13: <i>Post-hoc</i> definierte Subgruppen.....	69
Tabelle 4-14: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel .....	74
Tabelle 4-15: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	75
Tabelle 4-16: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern / Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel.....	77
Tabelle 4-17: Relevante Studien aus der Suche auf der Internetseite des G-BA – RCT mit dem zu bewertenden Arzneimittel.....	78
Tabelle 4-18: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	79
Tabelle 4-19: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel .....	81
Tabelle 4-20: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	83
Tabelle 4-21: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel.....	84

Tabelle 4-22: Nummerierung der einzelnen Items des <i>BPI-Fragebogens</i> für die Stratifizierung in der Studie UX023-CL303 .....	88
Tabelle 4-23: Auszug aus den Begleitmedikation während der doppelblinden Phase der Studie UX023-CL303 .....	92
Tabelle 4-24: Schmerzmittelgebrauch bei Studienvisite.....	93
Tabelle 4-25: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	95
Tabelle 4-26: Nummerierung der einzelnen Items des <i>BPI-Fragebogens</i> für die Stratifizierung in der Studie UX023-CL303 .....	96
Tabelle 4-27: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel .....	98
Tabelle 4-28: Operationalisierung von <i>Serumphosphat</i> .....	102
Tabelle 4-29: Bewertung des Verzerrungspotenzials für <i>Serumphosphat</i> in RCT mit dem zu bewertenden Arzneimittel .....	103
Tabelle 4-30: Ergebnisse für <i>Serumphosphat</i> aus RCT mit dem zu bewertenden Arzneimittel.....	105
Tabelle 4-31: Ergebnisse für <i>Serumphosphat</i> – Responderanalyse aus RCT mit dem zu bewertenden Arzneimittel (Woche 24) .....	107
Tabelle 4-32: Nummerierung der einzelnen Items des <i>BPI-Fragebogens</i> für die Stratifizierung in der Studie UX023-CL303 .....	108
Tabelle 4-33: Operationalisierung von <i>Gehfähigkeit</i> .....	110
Tabelle 4-34: Bewertung des Verzerrungspotenzials für <i>Gehfähigkeit</i> in RCT mit dem zu bewertenden Arzneimittel .....	112
Tabelle 4-35: Ergebnisse für <i>Gehfähigkeit</i> aus RCT mit dem zu bewertenden Arzneimittel (korrigierte Analyse) .....	113
Tabelle 4-36: Ergebnisse für <i>Gehfähigkeit</i> aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse).....	117
Tabelle 4-37: Operationalisierung von <i>Steifheit</i> und <i>physische Funktion</i> .....	120
Tabelle 4-38: Bewertung des Verzerrungspotenzials für <i>Steifheit und Physische Funktion</i> in RCT mit dem zu bewertenden Arzneimittel .....	122
Tabelle 4-39: Ergebnisse der <i>Steifheit</i> und <i>Physischen Funktion</i> aus RCT mit dem zu bewertenden Arzneimittel .....	124
Tabelle 4-40: Ergebnisse für <i>Steifheit und Physische Funktion</i> – Responderanalyse aus RCT mit dem zu bewertenden Arzneimittel (Woche 24) .....	128
Tabelle 4-41: Nummerierung der einzelnen Items des <i>BPI-Fragebogens</i> für die Stratifizierung in der Studie UX023-CL303 .....	129
Tabelle 4-42: Ergebnisse der <i>Steifheit</i> und <i>Physischen Funktion</i> aus RCT mit dem zu bewertenden Arzneimittel (MMRM-Analyse).....	131
Tabelle 4-43: Operationalisierung von <i>Schmerzerhebung</i> .....	135

Tabelle 4-44: Bewertung des Verzerrungspotenzials für <i>Schmerzerhebung</i> in RCT mit dem zu bewertenden Arzneimittel.....	137
Tabelle 4-45: Ergebnisse für <i>Schmerzerhebung (BPI Schlimmster Schmerz)</i> aus RCT mit dem zu bewertenden Arzneimittel.....	138
Tabelle 4-46: Nummerierung der einzelnen Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303 .....	140
Tabelle 4-47: Ergebnisse für <i>Schmerzerhebung – Responderanalyse</i> für <i>BPI Schlimmster Schmerz</i> aus RCT mit dem zu bewertenden Arzneimittel.....	142
Tabelle 4-48: Ergebnisse für <i>Schmerzerhebung (BPI Schmerzintensität und BPI Schmerzinterferenz)</i> aus RCT mit dem zu bewertenden Arzneimittel.....	143
Tabelle 4-49: Operationalisierung von <i>Fatigue</i> .....	147
Tabelle 4-50: Bewertung des Verzerrungspotenzials für <i>Fatigue</i> in RCT mit dem zu bewertenden Arzneimittel .....	148
Tabelle 4-51: Ergebnisse für <i>Fatigue</i> aus RCT mit dem zu bewertenden Arzneimittel .....	149
Tabelle 4-52: Operationalisierung von <i>Erfassung von Therapieeffekten</i> .....	152
Tabelle 4-53: Bewertung des Verzerrungspotenzials für <i>Erfassung von Therapieeffekten</i> in RCT mit dem zu bewertenden Arzneimittel .....	153
Tabelle 4-54: Ergebnisse für <i>Erfassung von Therapieeffekten</i> aus RCT mit dem zu bewertenden Arzneimittel .....	154
Tabelle 4-55: Operationalisierung von <i>Sicherheit</i> .....	157
Tabelle 4-56: Bewertung des Verzerrungspotenzials für <i>Sicherheit</i> in RCT mit dem zu bewertenden Arzneimittel .....	160
Tabelle 4-57: Ergebnisse für <i>unerwünschte Ereignisse nach Beginn der Behandlung (TEAE)</i> aus RCT mit dem zu bewertenden Arzneimittel ( <i>Safety-Population</i> ).....	161
Tabelle 4-58: Ergebnisse für <i>TEAE nach System Organ Class (SOC) und Preferred Term (PT)</i> aus RCT mit dem zu bewertenden Arzneimittel ( <i>Safety-Population</i> ).....	162
Tabelle 4-59: Ergebnisse für <i>TEAE von besonderem Interesse nach System Organ Class (SOC) und Preferred Term (PT)</i> aus RCT mit dem zu bewertenden Arzneimittel ( <i>Safety-Population</i> ) .....	165
Tabelle 4-60: Präspezifizierte Subgruppenanalysen .....	168
Tabelle 4-61: <i>Post-hoc</i> durchgeführte Subgruppenanalysen .....	169
Tabelle 4-62: p-Werte der Interaktion für präspezifizierte Subgruppenanalysen .....	170
Tabelle 4-63: p-Werte der Interaktion für <i>post-hoc</i> durchgeführte Subgruppenanalysen .....	173
Tabelle 4-64: Subgruppenanalysen – Ergebnisse für <i>Gehfähigkeit</i> aus RCT mit dem zu bewertenden Arzneimittel .....	177
Tabelle 4-65: Subgruppenanalysen – Ergebnisse für <i>Erfassung von Therapieeffekten</i> aus RCT mit dem zu bewertenden Arzneimittel .....	183
Tabelle 4-66: Subgruppenanalysen – Ergebnisse für <i>WOMAC Steifheit</i> aus RCT mit dem zu bewertenden Arzneimittel (Subgruppe <i>WOMAC Physische Funktion</i> ) .....	186

Tabelle 4-67: Subgruppenanalysen – Ergebnisse für <i>WOMAC Steifheit</i> aus RCT mit dem zu bewertenden Arzneimittel (Subgruppe <i>WOMAC Gesamtscore</i> ).....	187
Tabelle 4-68: Subgruppenanalysen – Ergebnisse für <i>WOMAC Steifheit</i> aus RCT mit dem zu bewertenden Arzneimittel (Subgruppe <i>6MWT</i> ).....	188
Tabelle 4-69: Liste der eingeschlossenen Studien - RCT .....	189
Tabelle 4-70: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche .....	190
Tabelle 4-71: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	191
Tabelle 4-72: Operationalisierung von <Endpunkt xxx>.....	191
Tabelle 4-73: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche .....	192
Tabelle 4-74: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche.....	192
Tabelle 4-75: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien .....	194
Tabelle 4-76: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien .....	195
Tabelle 4-77: Operationalisierung von <Endpunkt xxx>.....	195
Tabelle 4-78: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien .....	196
Tabelle 4-79: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen ....	198
Tabelle 4-80: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen.....	198
Tabelle 4-81 Zusammenfassung der Ergebnisse der Kategorie <i>Morbidität</i> zum medizinischen Zusatznutzen von Burosumab .....	202
Tabelle 4-82 Zusammenfassung der Ergebnisse der Kategorie <i>Sicherheit</i> zum medizinischen Zusatznutzen von Burosumab gegenüber der zVT .....	209
Tabelle 4-83: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	215
Tabelle 4-84: Dokumentation der Suchstrategie für randomisierte kontrollierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in EMBASE.....	224
Tabelle 4-85: Dokumentation der Suchstrategie für randomisierte kontrollierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in MEDLINE.....	225
Tabelle 4-86: Dokumentation der Suchstrategie für randomisierte kontrollierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in Cochrane.....	226
Tabelle 4-87: Dokumentation der Suchstrategie für randomisierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in clinicaltrials.gov .....	227
Tabelle 4-88: Dokumentation der Suchstrategie für randomisierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in EU Clinical Trial Register (EU-CTR).....	228

Tabelle 4-89: Dokumentation der Suchstrategie für randomisierte Studien mit Burosumab zur Behandlung der XLH (Crysvita®) in <i>WHO International Clinical Trials Registry Platform</i> (WHO ICTRP) .....	228
Tabelle 4-90: Zusammenfassende Darstellung der aus- und eingeschlossenen Studien .....	231
Tabelle 4-91: Ausgeschlossene Studien ClinicalTrials.gov .....	231
Tabelle 4-92: Ausgeschlossene Studien EU-CTR .....	234
Tabelle 4-93: Ausgeschlossene Studien ICTRP.....	234
Tabelle 4-94 (Anhang): Studiendesign und -methodik für Studie UX023-CL303 .....	239
Tabelle 4-95 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie UX023-CL303 .....	275
Tabelle 4-96: Liste der für die Bewertung herangezogenen Quellen.....	275
Tabelle 4-97: Nummerierung der einzelnen Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303 .....	276
Tabelle 4-98: Nummerierung der Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303 .....	278

**Abbildungsverzeichnis**

	<b>Seite</b>
Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel .....	76
Abbildung 2: Studiendesign der Studie UX023-CL303 .....	88
Abbildung 3: Mittelwerte für die Endpunkte <i>WOMAC Steifheit</i> und <i>WOMAC Physische Funktion</i> über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303.....	126
Abbildung 4: Mittelwerte für den Endpunkt <i>BPI Schlimmster Schmerz</i> über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303 .....	139
Abbildung 5: Mittelwerte für den Endpunkte <i>BPI Schmerzintensität</i> und <i>BPI Schmerzinterferenz</i> über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303.....	145
Abbildung 6: Mittelwerte für den Endpunkt <i>Fatigue (BFI Schlimmste Fatigue und BFI Globaler Fatigue)</i> über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303.....	150
Abbildung 7: Mittelwerte für den Endpunkt <i>Erfassung von Therapieeffekten</i> über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303.....	155
Abbildung 8: Subgruppenanalyse <i>PGI-C</i> .....	185
Abbildung 9: Patientenfluss der Studie UX023-CL303 gemäß CONSORT .....	273

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
6MWT	6-Minuten-Gehdistanz Test
AESI	A priori definierte UE von besonderem Interesse ( <i>engl.: adverse event of special interest</i> )
AMIS	Arzneimittel-Informationssystem
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ARD	Absolute Risikodifferenz ( <i>engl.: Absolute Risk Difference</i> )
ARR	Adjustiertes relatives Risiko ( <i>engl.: Adjusted Risk Ratio</i> )
ATS	<i>American Thoracic Society</i>
BFI	<i>Brief Fatigue Inventory</i>
BMI	Body-Mass-Index
BOCF	<i>Baseline Observation Carried Forward</i>
B-P	Behandlungsdifferenz betrachtet von Baseline zu Post-Baseline
BPI	<i>Brief Pain Inventory</i>
CMH	Cochran–Mantel–Haenszel-Test
COMP	<i>Committee for Orphan Medicinal Products</i>
CONSORT	<i>Consolidated Standards of Reporting Trials</i>
CRF	<i>Case Report Form</i>
CTCAE	<i>Common Terminology Criteria for Adverse Events</i>
DGKED	Deutschen Gesellschaft für Kinderendokrinologie und -diabetologie
EDC	<i>Electronic Data Capture System</i>
EG	Europäische Gemeinschaft
EMA	Europäische Arzneimittelagentur
EU-CTR	<i>EU Clinical Trials Registry</i>
FAS	<i>Full Analysis Set</i>
FDA	<i>Food and Drug Administration</i>
FGF23	Fibroblasten-Wachstumsfaktor 23
FWER	<i>Family-Wise-Error-Rate</i>
G-BA	Gemeinsamer Bundesausschuss
GEE	<i>Generalized estimation equation</i>
H'g	<i>Hedges' g</i>
ICC	<i>Intraclass coefficients</i>

<b>Abkürzung</b>	<b>Bedeutung</b>
IgG1	Immunglobulin G1
ITT	<i>Intention-To-Treat</i>
IWRS	<i>Interactive web response system</i>
KG	Körpergewicht
KI	Konfidenzintervall
LLN	<i>Lower limit of normal</i>
LOCF	<i>Last Observation Carried Forward</i>
LS Mean	<i>Least Squares Means</i>
mBOCF	<i>Modified Baseline Observation Carried Forward</i>
MedDRA	<i>Medical Dictionary for Regulatory Activities</i>
MeSH	<i>Medical Subject Headings</i>
MID	Klinische Relevanzschwelle ( <i>engl.: minimally important difference</i> )
MMRM	<i>Mixed effect Model Repeat Measurement</i>
MTC	<i>Mixed Treatment Comparison</i>
MW	Mittelwert
MWD	Mittelwertdifferenz
NaPi	Natrium-Phosphat-Kotransportersystem
nu	Normierte Einheiten ( <i>engl.: normalised units</i> )
OMERACT	<i>Outcome Measures in Rheumatology Arthritis Clinical Trials</i>
OR	<i>Odds Ratio</i>
PGI-I	<i>Patient Global Impression of Improvement</i>
PGI-S	<i>Patient Global Impression of Severity</i>
PHEX	<i>Phosphat regulierendes Gen mit Homologie zu Endopeptidasen auf dem X-Chromosom (engl.: Phosphate-regulating neutral endopeptidase)</i>
PT	<i>Preferred Terms nach MedDRA</i>
RCT	<i>Randomized Controlled Trial</i>
RR	Relatives Risiko ( <i>engl.: Risk Ratio</i> )
SD	Standardabweichung
SE	Standardfehler
SF-36	<i>Short-Form 36 Health Survey</i>
SGB	Sozialgesetzbuch

<b>Abkürzung</b>	<b>Bedeutung</b>
SMQs	<i>Standardised MedDRA Queries</i>
SOC	<i>System Organ Class</i> nach MedDRA
STE	<i>Surrogate Threshold Effects</i>
STROBE	<i>Strengthening the Reporting of Observational Studies in Epidemiology</i>
SUE	Schwerwiegendes UE
TEAE	Unerwünschtes Ereignis nach Behandlungsbeginn ( <i>engl.: treatment emergent adverse event</i> )
TMP/GFR	Tubuläres Maximum der Phosphatrückresorption
TREND	<i>Transparent Reporting of Evaluations with Non-Randomized Design</i>
UE	Unerwünschtes Ereignis
VerfO	Verfahrensordnung
WHO	<i>World Health Organization</i>
WHO ICTRP	<i>WHO International Clinical Trials Registry Platform</i>
WOMAC	<i>Western Ontario and McMaster Universities Osteoarthritis</i>
XLH	X-chromosomale Hypophosphatämie
zVT	Zweckmäßige Vergleichstherapie

#### 4 Modul 4 – allgemeine Informationen

Modul 4 enthält folgende Angaben:

- Zusammenfassung (Abschnitt 4.1)
- Angaben zur Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens (Abschnitt 4.2)
- Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen (Abschnitt 4.3)
- eine abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens, einschließlich der Angabe von Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht (Abschnitt 4.1)
- ergänzende Informationen zur Begründung der vorgelegten Unterlagen (Abschnitt 4.5)

Für jedes zu bewertende Anwendungsgebiet ist eine separate Version des vorliegenden Dokuments zu erstellen. Die Kodierung der Anwendungsgebiete ist in Modul 2 hinterlegt. Sie ist je Anwendungsgebiet einheitlich für die Module 3, 4 und 5 zu verwenden.

Im Dokument verwendete Abkürzungen sind in das Abkürzungsverzeichnis aufzunehmen. Sofern Sie für Ihre Ausführungen Tabellen und Abbildungen verwenden, sind diese im Tabellen- bzw. Abbildungsverzeichnis aufzuführen.

## 4.1 Zusammenfassung der Inhalte von Modul 4

*Stellen Sie eine strukturierte Zusammenfassung der Inhalte von Modul 4 zur Verfügung.*

### Hintergrund

Bei Burosumab handelt es sich um ein Arzneimittel zur Behandlung eines seltenen Leidens. Für Arzneimittel zur Behandlung eines seltenen Leidens gilt der Zusatznutzen durch die Zulassung als belegt. Nachweise gemäß § 35a Abs. 1 Satz 3 Nummer 2 und 3 SGB V über den medizinischen Nutzen und Zusatznutzen im Verhältnis zur zweckmäßigen Vergleichstherapie (zVT) müssen nicht erbracht werden. Lediglich das Ausmaß des Zusatznutzens ist für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, nachzuweisen (5. Kapitel § 12 Nr. 1 Satz 2 VerfO des G-BA). Übersteigt der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten zwölf Kalendermonaten einen Betrag von 50 Millionen Euro, hat der pharmazeutische Unternehmer gemäß § 35a Absatz 1 Satz 12 SGB V innerhalb von drei Monaten nach Aufforderung durch den G-BA Nachweise gemäß 5. Kapitel § 5 Absatz 1 bis 6 VerfO des G-BA zu übermitteln.

Der G-BA hat mit Schreiben vom 21.10.2021 die Kyowa Kirin GmbH darüber informiert, dass der Umsatz mit Burosumab (Crysvita®) zu Lasten der GKV einen Betrag von 50 Millionen Euro überschritten hat. Maßgeblicher Zeitpunkt für die Einreichung der Nachweise gemäß § 35a Abs. 1 Satz 12 SGB V ist damit der 01.02.2022. Im vorliegenden Dossier werden nun die erforderlichen Nachweise des Zusatznutzens von Burosumab gegenüber der zVT dargestellt.

Für die erneute Bewertung von Burosumab nach Überschreitung der Umsatzgrenze von 50 Millionen Euro werden die Studiendaten der Studie UX023-CL303, eine randomisierte, doppelblinde, Placebo-kontrollierte, multizentrische Studie der Phase III (24 Wochen) mit einer Evidenzstufe Ib (RCT) herangezogen, die bereits in dem Nutzenbewertungsverfahren 2020-11-01-D-588 bewertet wurde [1].

Laut Anwendungsgebiet ist Burosumab zur Behandlung der XLH bei Kindern und Jugendlichen im Alter von 1 – 17 Jahren mit röntgenologischem Nachweis einer Knochenerkrankung, sowie bei Erwachsenen angezeigt [2]. Das Modul 4 B bezieht sich jedoch ausschließlich auf Erwachsene mit XLH.

### Fragestellung

Wie groß ist das Ausmaß des Zusatznutzens von Burosumab (Crysvita®) bezüglich patientenrelevanter Endpunkte bei der Behandlung von Erwachsenen mit X-chromosomaler Hypophosphatämie (XLH), die unter einer Phosphatsubstitution innerhalb eines Jahres keine Verbesserung der klinischen Symptome zeigen oder bei denen eine Phosphatsubstitution kontraindiziert ist?

Für die Nutzenbewertung ist zu beachten, dass zwischen zwei Teilpopulationen unterschieden wird, die sich hinsichtlich des Ansprechens auf eine Substitutionstherapie aus oralem Phosphat und aktivem Vitamin D unterscheiden:

### 1. Teilpopulation

Erwachsene, die innerhalb **eines Jahres** auf eine Phosphatsubstitution **ansprechen**

### 2. Teilpopulation

Erwachsene, die innerhalb **eines Jahres nicht** auf eine Phosphatsubstitution **ansprechen**

Trotz langjähriger Phosphatsubstitution wird bei fast allen Erwachsenen mit XLH das **primäre Therapieziel**, eine **Serumphosphat-Konzentration im unteren Normbereich** (*lower limit of normal* (LLN):  $\geq 2,5$  mg/dl), nicht erlangt. Aufgrund des bestehenden chronischen Phosphatverlusts treten schwerwiegende Folgeerkrankungen wie Skelettdeformationen, Arthrose, Osteomalazie, Pseudofrakturen, Enthesopathien und Spinalkanalstenosen auf, die mit starken Schmerzen und erheblichen Einschränkungen im Alltag einhergehen. Des Weiteren führt die Phosphatsubstitution zu schweren Nebenwirkungen wie z. B. zu Nephrokalzinosen, Nephrolithiasis und sekundärem oder tertiärem Hyperparathyreoidismus. Aufgrund dessen wird die Therapie häufig im Erwachsenenalter abgebrochen.

Bis zum heutigen Tag existieren keine Leitlinien für Erwachsene mit XLH, die eine Phosphatsubstitution empfehlen. Zudem liegen keine Endpunktstudien mit ausreichender Evidenz vor, die den Nutzen einer Phosphatsubstitution bei erwachsenen XLH-Patienten belegen. Lediglich für Kinder wird laut der einzig verfügbaren S1-Leitlinie eine Phosphatsubstitution empfohlen [3]. Da für Erwachsene mit XLH keine den Pathomechanismus adressierende Therapieoption zur Verfügung steht, werden diese anschließend neben teils schweren Operationen rein symptomatisch mit einer Schmerztherapie behandelt.

Das unzureichende Ansprechen wird auch anhand der Charakteristika der Studienteilnehmer der Studie UX023-CL303 ersichtlich. Trotz jahrelanger Phosphatsubstitution weisen alle Studienteilnehmer Folgeerkrankungen der XLH auf [4, 5]. Zudem benötigten bereits zwei Drittel der XLH-Patienten im Kindes- und Jugendalter orthopädische Eingriffe, um die stark ausgeprägten Skelettdeformationen zu beheben. Bereits 35 % der 18 – 29-Jährigen leiden frühzeitig unter Arthrose, wobei die Gesamtrate in der deutschen Gesamtbevölkerung lediglich 1,7 % beträgt [6]. Bei den 30 – 39-Jährigen sind es sogar 74 %, in der Gesamtbevölkerung dagegen nur 6,7 %.

Da keine den Pathomechanismus adressierende Therapieoption verfügbar ist, kann für diese Teilpopulation lediglich „*best supportive care*“ als zVT im Anwendungsgebiet angesehen werden.

Anhand der Charakteristika der Studienteilnehmer kann davon ausgegangen werden, dass nahezu alle behandlungsbedürftigen Erwachsenen mit XLH der Teilpopulation zuzuordnen sind, die nicht auf eine Phosphatgabe ansprechen. Aufgrund dessen wird der Zusatznutzen von Burosumab in dem vorliegenden Dossier ausschließlich für diese Teilpopulation abgeleitet.

### **Datenquellen**

Zur Bewertung des Zusatznutzens von Burosumab wird die Zulassungsstudie UX023-CL303 herangezogen. Hierbei handelt es sich um eine randomisierte, doppelblinde, Placebo-

kontrollierte multizentrische Studie der Phase III, die die Wirksamkeit und Sicherheit von Burosumab im Vergleich zu Placebo untersucht. In die Studie wurden Erwachsene mit XLH ( $\geq 18 - 65$  Jahre) eingeschlossen und über einen Zeitraum von bis zu 149 Wochen untersucht. Aufgrund des offenen Studiendesigns nach Woche 24 wurde zum Ableiten des Zusatznutzens jedoch ausschließlich die 24-wöchige doppelblinde Phase der Studie berücksichtigt. Die Studienteilnehmer wurden von Studienzentren in den USA und den europäischen Ländern Frankreich, Italien, Irland und Großbritannien rekrutiert.

### **Ein-/Ausschlusskriterien für Studien**

#### ***Einschlusskriterien für Studien***

Für die Nutzenbewertung von Arzneimitteln sollen ausschließlich Studien mit einer hohen Evidenzstufe und einem niedrigem Verzerrungspotenzial eingeschlossen werden. Daher wurden lediglich Studien berücksichtigt, deren Studiendesign diesen Kriterien entsprach. Hierzu zählen randomisierte kontrollierte Studien, die in dem Anwendungsgebiet Erwachsene mit XLH mit einer Mindestlaufzeit von 24 Wochen durchgeführt wurden. Eine ausführliche Beschreibung ist in Tabelle 4-1 zu finden.

Tabelle 4-1: Einschlusskriterien der Studien für die Nutzenbewertung von Burosumab

Einschlusskriterien		
<b>E1</b>	<b>Patientenpopulation</b>	Erwachsene ( $\geq 18$ Jahre) mit X-chromosomaler Hypophosphatämie (XLH) <sup>a</sup>
<b>E2</b>	<b>Intervention</b>	Burosumab, Crysvisa <sup>®</sup> , KRN23 1,0 mg/kg KG alle 4 Wochen Höchstdosis: 90 mg Subkutane Applikation
<b>E3</b>	<b>Komparator</b>	Placebo + Begleittherapie (= <i>best supportive care</i> ) <sup>b</sup>
<b>E4</b>	<b>Endpunkte</b>	Erhebung von mindestens einem patientenrelevanten Endpunkt, wie zum Beispiel Morbidität (z. B. Schmerz, Fatigue, Steifheit), Lebensqualität und unerwünschte Ereignisse
<b>E5</b>	<b>Studiendesign</b>	Randomisierte, kontrollierte Studien (RCT) Doppelblind
<b>E6</b>	<b>Studiendauer</b>	$\geq 24$ Wochen
<b>E7</b>	<b>Publikationstyp</b>	Volltextpublikation oder Studienbericht Veröffentlichung in deutscher oder englischer Sprache

a: Anhand der Literatur und den Charakteristika der Studienteilnehmer der Studie UX023-CL303 kann davon ausgegangen werden, dass nahezu alle erwachsenen XLH-Patienten der Teilpopulation *Erwachsene, die innerhalb eines Jahres nicht auf eine Phosphatsubstitution ansprechen* zuzuordnen sind. Eine Einschränkung bei der Suche nach Studien im Anwendungsgebiet ist daher nicht notwendig.

b: Der G-BA hat für das vorliegende Anwendungsgebiet als zVT eine Phosphatsubstitution bestimmt [7]. Nach Auffassung der Kyowa Kirin GmbH stellt die Phosphatsubstitution für Erwachsene mit XLH jedoch eine unzureichende Therapie dar. Da keine weiteren Therapieoptionen zur Verfügung stehen, kommt für erwachsene XLH-Patienten lediglich „*best supportive care*“ als zVT in Betracht. Als geeignete Umsetzung von „*best supportive care*“ wird Placebo in Kombination mit einer Begleittherapie gesehen. Eine ausführliche Begründung ist in Modul 3 Abschnitt 3.1.2 aufgeführt. Um die formale Vollständigkeit zu gewährleisten, wurde dennoch eine Suche nach Studien mit dem Komparator Phosphat durchgeführt. Es konnten jedoch keine Studien identifiziert werden, die diesen historisch gewachsenen Therapieansatz verwenden und gleichzeitig die in der Tabelle aufgeführten Einschlusskriterien (E1 – E2) und E4 – E7) erfüllten.

### Ausschlusskriterien für Studien

Nicht-randomisierte kontrollierte Studien sowie Studien mit einem anderen Wirkstoff oder einer nicht zulassungskonformen Dossierung und Applikation von Burosumab wurden in dem vorliegenden Dossier nicht berücksichtigt. Des Weiteren wurden Studien, deren Studienlaufzeit kürzer als 24 Wochen war oder von denen keine Vollpublikationen sowie Studienregistereinträge vorlagen, ausgeschlossen. Mehrfachpublikationen von Studien wurden ebenfalls nicht mit einbezogen. Eine ausführliche Beschreibung der Ausschlusskriterien ist in Tabelle 4-2 zu finden.

Tabelle 4-2: Ausschlusskriterien der Studien für die Nutzenbewertung von Burosumab

Ausschlusskriterien		
A1	Patientenpopulation nicht entsprechend E1	Patienten mit XLH < 18 Jahren
A2	Intervention nicht entsprechend E2	Andere Wirkstoffe Dosierung, Applikation nicht gemäß Zulassung von Burosumab
A3	Komparator	Andere Wirkstoffe
A4	Endpunkte nicht entsprechend E4	Studien, bei denen ausschließlich Surrogat-Parameter evaluiert werden
A5	Studiendesign nicht entsprechend E5	Nicht-randomisierte, kontrollierte Studien
A6	Studiendauer nicht entsprechend E6	< 24 Wochen
A7	Publikationstyp nicht entsprechend E7	Keine Primärpublikation, Review, Abstract, Kongress-Poster Andere Sprache als Deutsch oder Englisch

### Methoden zur Bewertung der Aussagekraft der Nachweise und zur Synthese von Ergebnissen

Zur Bewertung der Aussagekraft der Nachweise wurde sowohl das Verzerrungspotenzial auf Studienebene als auch auf Endpunktebene evaluiert.

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde, Placebo-kontrollierte, multizentrische Studie der Phase III, die, entsprechend der Kriterien des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), der Evidenzstufe 1b entspricht [8]. Sowohl den Studienteilnehmern, dem Studienpersonal als auch dem Prüfarzt war die Gruppenzuteilung während der 24-wöchigen, doppelblinden Phase der Studie nicht bekannt. Die patientenberichteten Fragebögen wurden von den Studienteilnehmern selbst ausgefüllt, und biochemische Analysen erfolgten objektiv und verblindet in einem zentralen Labor.

Aufgrund der fehlerhaften Stratifizierung zu Studienbeginn (Abschnitt 4.2.5.4, Anhang 4-F) wurden sowohl für den primären Wirksamkeitsendpunkt (*Serumphosphat-Prozentsatz erreichter Normwert*) als auch für die sekundären Wirksamkeitsendpunkte (*Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Steifheit, WOMAC Physische Funktion, Brief Pain Inventory (BPI) Schlimmster Schmerz*) Sensitivitätsanalysen durchgeführt, um die Robustheit der Ergebnisse zu überprüfen. Diese bestätigten die Ergebnisse der Primäranalysen, sodass weiterhin von einer hohen Ergebnissicherheit ausgegangen werden kann. Somit liegt sowohl auf Studienebene als auch auf Endpunktebene während der gesamten doppelblinden Phase der Studie (24 Wochen) ein niedriges Verzerrungspotenzial vor. Details hierzu sind sowohl im Abschnitt 4.3.1.2.2 (Verzerrungspotenzial auf Studienebene) als auch im Abschnitt 4.3.1.3 (Verzerrungspotenzial auf Endpunktebene) zu finden.

Für die Synthese der Ergebnisse wurde der Studienbericht der Studie UX023-CL303 sowie dazugehörige *post-hoc* durchgeführte Zusatzanalysen herangezogen.

### Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen

Die Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen beruhen auf der Studie UX023-CL303, einer randomisierten, doppelblinden, Placebo-kontrollierten multizentrischen Studie der Phase III. Für die Herleitung des Zusatznutzens wurde die doppelblinde Phase (24 Wochen) der Studie berücksichtigt. Da sowohl auf Studienebene als auch auf Endpunktebene ein niedriges Verzerrungspotenzial vorliegt, kann aus der Studie ein **Hinweis** auf einen Zusatznutzen abgeleitet werden. Eine Übersicht der Ergebnisse ist in der Tabelle 4-3 und in der Tabelle 4-4 für jeden Endpunkt aufgeführt.

#### Mortalität

Bei der XLH handelt es sich um eine genetisch bedingte schwere chronische Erkrankung, die grundsätzlich nicht potenziell tödlich verläuft. Übereinstimmend hiermit konnte während der gesamten Studiendauer kein Todesfall festgestellt werden. Daher wurde in der Studie UX023-CL303 die Mortalität anhand der schweren unerwünschten Ereignisse erfasst.

#### Morbidität

Die Tabelle 4-3 enthält eine Übersicht der Ergebnisse zu patientenrelevanten Endpunkten der Kategorie *Morbidität*.

Tabelle 4-3: Zusammenfassung der Ergebnisse der Kategorie *Morbidität* zum medizinischen Zusatznutzen von Burosumab gegenüber der zweckmäßigen Vergleichstherapie (zVT)

Patientenrelevante Endpunkte	UX023-CL303 Burosumab (N = 68) vs. Placebo (N = 66) Behandlungsdifferenz (B-P) Burosumab vs. Placebo Woche 24
<b>Serumphosphat</b>	
<i>Serumphosphat (mg/dl)</i>	
LsMean <sup>a</sup>	0,40
[95% KI]	[0,28; 0,51]
p-Wert	< 0,0001
Hedges`g <sup>b</sup> [95% KI]	H`g = 0,922 [0,566; 1,278]
<i>Serumphosphat Prozentsatz erreichter Normwert (≥ 2,5 mg/dl)</i>	
Burosumab (n (%)) vs. Placebo (n (%))	63 (92,6) vs. 5 (7,6)
RR <sup>a</sup>	12,229
[95% KI]	[5,252; 28,479]
p-Wert	< 0,0001
<b>Gefähigkeit</b>	
<i>6MWT – Messung in Meter<sup>c</sup></i>	
LsMean <sup>a</sup>	19,83
[95% KI]	[4,19; 35,47]
p-Wert	0,0129
Hedges`g <sup>b</sup> [95% KI]	H`g = 0,424 [0,075; 0,773]

<b>Patientenrelevante Endpunkte</b>	<b>UX023-CL303 Burosumab (N = 68) vs. Placebo (N = 66) Behandlungsdifferenz (B-P) Burosumab vs. Placebo Woche 24</b>
<i>6MWT Gesamtrate – Prozentsatz<sup>c</sup></i>	
LsMean <sup>a</sup>	3,19
[95% KI]	[0,96; 5,41]
p-Wert	0,0050
Hedges`g <sup>b</sup> [95% KI]	H`g = 0,460 [0,110; 0,809]
<b>Steifheit und Physische Funktion</b>	
<i>WOMAC Steifheit</i>	
LsMean <sup>a</sup>	-8,31
[95% KI]	[-14,68, -1,94]
p-Wert	0,0106
Hedges`g <sup>b</sup> [95% KI]	H`g = -0,420 [-0,765; -0,075]
<i>WOMAC Steifheit ≥ 15 nu</i>	
Burosumab (n (%)) vs. Placebo (n (%))	21 (30,9) vs. 10 (15,2)
RR <sup>a</sup>	2,038
[95% KI]	[1,040; 3,994]
p-Wert	0,0403
<i>WOMAC Physische Funktion</i>	
LsMean <sup>a</sup>	-4,90
[95% KI]	[-9,76, -0,05]
p-Wert	0,0478
Hedges`g <sup>b</sup> [95% KI]	H`g = -0,389 [-0,735; -0,043]
<i>WOMAC Physische Funktion ≥ 15 nu</i>	
Burosumab (n (%)) vs. Placebo (n (%))	14 (20,6) vs. 11 (16,7)
RR <sup>a</sup>	1,235
[95% KI]	[0,605; 2,521]
p-Wert	0,6591
<b>Schmerzerhebung</b>	
<i>BPI Schlimmster Schmerz</i>	
LsMean <sup>a</sup>	-0,46
[95% KI]	[-1,00; 0,08]
p-Wert	0,0919
Hedges`g <sup>b</sup> [95% KI]	-
<i>BPI Schlimmster Schmerz ≥ 15 %</i>	
Burosumab (n (%)) vs. Placebo (n (%))	29 (42,6) vs. 23 (34,8)
RR <sup>a</sup>	1,224
[95% KI]	[0,796; 1,881]
p-Wert	0,3800

<b>Patientenrelevante Endpunkte</b>	<b>UX023-CL303 Burosumab (N = 68) vs. Placebo (N = 66) Behandlungsdifferenz (B-P) Burosumab vs. Placebo Woche 24</b>
<i>BPI Schlimmster Schmerz <math>\geq 30</math> %</i> Burosumab (n (%)) vs. Placebo (n (%)) RR <sup>a</sup> [95% KI] p-Wert	17 (25,0) vs. 12 (18,2) 1,375 [0,713; 2,652] 0,4036
<i>BPI Schmerzintensität</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,43 [-0,93; 0,07] 0,0940 -
<i>BPI Schmerzinterferenz</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,13 [-0,70; 0,44] 0,6527 -
<b>Fatigue</b>	
<i>BFI Schlimmste Fatigue</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,20 [0,40; -0,80] 0,5102 -
<i>BFI Globale Fatigue</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	0,13 [0,70, -0,44] 0,6572 -
<b>Erfassung von Therapieeffekten</b>	
<i>PGI-I</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,29 [-0,74; 0,16] 0,2073 -
<p>(B-P): Behandlungsdifferenz betrachtet von Baseline zu Post-Baseline (Woche 24) der Burosumab-Gruppe mit Placebo; 6MWT: 6-Minuten-Gehdistanz Test; WOMAC: <i>Western Ontario and McMaster Universities Osteoarthritis</i>; BPI: <i>Brief Pain Inventory</i>; BFI: <i>Brief Fatigue Inventory</i>; PGI-I: <i>Patient Global Impression of Improvement</i>; [95% KI]:95%-Konfidenzintervall; RR: <i>Risk Ratio</i> (relatives Risiko)</p> <p>Ein RR bzw. OR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &lt; 1 einen Vorteil für Placebo.</p> <p>a: LsMean/RR: Eine detaillierte Beschreibung des statistischen Modells ist in den Abschnitten 4.2.5.2 und 4.3.1.3 zu finden.</p> <p>b: Hedges`g-Berechnung erfolgte nur falls <math>p &lt; 0,05</math>.</p> <p>c: 6MWT: Aufgrund eines Kalkulationsfehlers durch das Studienzentrum zu Beginn der Studie, wurde zunächst ein fehlerhafter Baseline-Wert in die Datenbank eingetragen, der nach der Primäranalyse zu Woche 24 korrigiert wurde. Die korrigierte Analyse berücksichtigt ausschließlich die tatsächlich zurückgelegte Distanz des Studienteilnehmers nach der</p>	

<b>Patientenrelevante Endpunkte</b>	<b>UX023-CL303</b> <b>Burosumab (N = 68) vs. Placebo (N = 66)</b> <b>Behandlungsdifferenz (B-P)</b> <b>Burosumab vs. Placebo</b> <b>Woche 24</b>
Korrektur des Kalkulationsfehlers. Diese korrigierte Analyse wurde bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 vom G-BA anerkannt und zur Ableitung des Zusatznutzens herangezogen [1].	

## ***Serumphosphat***

### *Serumphosphat (mg/dl)*

Burosumab stellt die erste wirksame und zielgerichtete Therapieoption dar, die es ermöglicht innerhalb einer kurzen Zeit die Serumphosphat-Homöostase (LLN:  $\geq 2,5$  mg/dl) von XLH-Patienten wiederherzustellen. Bereits nach 12 Wochen erreicht die Burosumab-Gruppe eine durchschnittliche *Serumphosphat-Konzentration* von 2,69 mg/dl, während die durchschnittliche *Serumphosphat-Konzentration* in der Placebo-Gruppe (2,06 mg/dl) weiterhin deutlich unter dem Normbereich liegt. Der positive Effekt auf die Serumphosphat-Homöostase bleibt auch langfristig bestehen. Die Behandlungsdifferenz zu Woche 24 belegt hierbei sowohl den statistisch signifikanten als auch klinisch relevanten Vorteil der Burosumab-Behandlung gegenüber der zVT (LsMean<sub>W24</sub>: 0,40 (mg/dl);  $p < 0,0001$ , H<sub>g</sub> = 0,922 [0,566; 1,278]) (Tabelle 4-3).

### *Prozentsatz erreichter Normwert (LLN: $\geq 2,5$ mg/dl)*

Anhand des Anteils an Studienteilnehmern mit einer als klinisch relevant eingestuften Verbesserung der *Serumphosphat-Konzentration* (LLN:  $\geq 2,5$  mg/dl) zeigt sich zu Woche 24 eine statistische Signifikanz zugunsten von Burosumab (Burosumab: 92,6 %; Placebo: 7,6 %;  $p < 0,0001$ ) (Tabelle 4-3).

## ***Gehfähigkeit***

### *6MWT (absoluter und relativer Wert)*

Sowohl zu Woche 12 als auch zu Woche 24 zeigt sich anhand des *6MWT* (absoluter Wert: Gesamtdistanz in Meter, relativer Wert: Prozentsatz des vorhergesagten Werts auf Grundlage normativer Daten) eine deutliche Steigerung der Gehfähigkeit in der Burosumab-Gruppe, während sich die Placebo-Gruppe im Vergleich zu Baseline verschlechtert. Die statistisch signifikante Behandlungsdifferenz (absoluter 6MWT: LsMean: 19,83 m;  $p = 0,0129$ ; relativer 6MWT: LsMean: 3,19 Prozentpunkte;  $p = 0,005$ ) unterstreicht den positiven Effekt von Burosumab auf die Mobilität der XLH-Patienten (Tabelle 4-3).

## ***Steifheit und Physische Funktion:***

### *WOMAC Steifheit*

Für die Domäne *WOMAC Steifheit* zeigt sich zu Woche 24 eine statistisch signifikante Behandlungsdifferenz (LsMean: -8,31;  $p = 0,0106$ ) zugunsten von Burosumab (Tabelle 4-3). Der Unterschied hinsichtlich des Anteils an Studienteilnehmern mit einer als klinisch relevant eingestuften Verbesserung der *Steifheit* (Abnahme um  $\geq 15$  nu) erreicht zu Woche 24 ebenfalls

eine statistische Signifikanz zugunsten von Burosumab (Burosumab: 30,9 %; Placebo: 15,2 %;  $p = 0,0403$ ) (Tabelle 4-3).

#### *WOMAC Physische Funktion*

Für die Domäne *WOMAC Physische Funktion* liegt ebenfalls zu Woche 24 ein statistisch signifikanter Effekt zugunsten von Burosumab vor (LsMean: -4,90;  $p = 0,0478$ ) (Tabelle 4-3). Auch wenn der Anteil an Studienteilnehmern mit einer als klinisch relevant eingestuften Verbesserung ( $\geq 15$  nu) zu Woche 24 keinen statistisch signifikanten Unterschied zwischen den Behandlungsgruppen aufweist, ist ein Trend zu einem größeren Anteil an Respondern in der Burosumab-Gruppe feststellbar (Burosumab: 20,6 %; Placebo: 16,7 %) (Tabelle 4-3).

#### **Schmerzerhebung**

##### *BPI Schlimmster Schmerz*

Zu Beginn der Studie wiesen die Studienteilnehmer durchschnittlich einen *BPI Schlimmsten Schmerz* von 6,8 auf einer Skala von 0 – 10 auf (Mild: 1 – 3, Moderat: 4 – 6, Stark: 7 – 10), wodurch sich das Ausmaß der starken chronischen Schmerzen, die XLH-Patienten aufgrund der schwerwiegenden Folgeerkrankungen ausgesetzt sind, zeigt. Trotz fehlender statistischer Signifikanz zeigt sich bereits zu Woche 24 eine deutliche Tendenz zu einer größeren Reduktion des *BPI Schlimmsten Schmerzes* zugunsten von Burosumab (LsMean: -0,46;  $p = 0,0919$ ) (Tabelle 4-3). Dies wird auch anhand des Anteils an Respondern deutlich, die eine als klinisch relevant eingestufte Verbesserung der Domäne *BPI Schlimmster Schmerz* zu Woche 24 erreichen. Während in der Burosumab-Gruppe 42,6 % der Studienteilnehmer eine Abnahme um mindestens 15 % (kleinster relevanter Unterschied) aufweisen, ist dies nur bei 34,8 % der Studienteilnehmer in der Placebo-Gruppe der Fall (Tabelle 4-3). Übereinstimmend hiermit ist ebenfalls der Anteil an Respondern, der eine bedeutsame Linderung (Abnahme um  $\geq 30$  %) der chronischen Schmerzen zu Woche 24 erreicht, in der Burosumab-Gruppe größer (Burosumab: 25 %; Placebo: 18,2 %) (Tabelle 4-3).

##### *BPI Schmerzintensität*

Für die Domäne *BPI Schmerzintensität* zeigt sich zu Woche 24 ein deutlicher Trend zu einer stärkeren Abnahme von XLH-bedingten Schmerzen unter der Burosumab-Therapie (LsMean: -0,43;  $p = 0,0940$ ) (Tabelle 4-3).

##### *BPI Schmerzinterferenz*

Für die Domäne *BPI Schmerzinterferenz* zeigt sich zu Woche 24 ebenfalls eine leichte Verbesserung zugunsten von Burosumab (LsMean: -0,13;  $p = 0,6527$ ) (Tabelle 4-3).

#### **Fatigue**

##### *BFI Schlimmste Fatigue*

Sowohl die Burosumab- als auch die Placebo-Gruppe weisen am Ende der doppelblinden Phase ein ähnliches Ausmaß an Fatigue auf (Behandlungsdifferenz Woche 24: LsMean: -0,2;  $p = 0,5102$ ) (Tabelle 4-3).

*BFI Globale Fatigue*

Für die Domäne *BFI Globale Fatigue* kann zu Woche 24 kein Unterschied (LsMean: 0,13;  $p = 0,6572$ ) zwischen den Behandlungsgruppen festgestellt werden (Tabelle 4-3).

*Erfassung von Therapieeffekten**PGI-I*

Anhand des *PGI-I* zeigt sich zu Woche 24 kein Behandlungsunterschied zwischen der Burosumab- und Placebo-Gruppe (LsMean: -0,29;  $p = 0,2073$ ) (Tabelle 4-3).

Zusammenfassend lässt sich feststellen, dass mit **Burosumab** erstmalig eine **zielgerichtete Behandlungsoption** zur Verfügung steht, die bereits nach **kurzer Zeit** (12 Wochen) die **Serumphosphat-Homöostase** bei Erwachsenen XLH-Patienten **wiederherstellt** und **langanhaltend** aufrechterhält (24 Wochen). Infolgedessen werden die Gehfähigkeit der XLH-Patienten signifikant gesteigert und das Ausmaß der Steifheit sowie der Einschränkung der physischen Funktion signifikant reduziert.

*Sicherheit*

Die Tabelle 4-4 enthält eine Übersicht der Ergebnisse der patientenrelevanten Endpunkte der Kategorie *Sicherheit* in der Studie UX023-CL303. Zur Ableitung des Zusatznutzens wurde ausschließlich die doppelblinde Placebo-kontrollierte Phase der Studie herangezogen.

Tabelle 4-4: Zusammenfassung der Ergebnisse der Kategorie *Sicherheit* zum medizinischen Zusatznutzen von Burosumab gegenüber der zVT

Sicherheit	Burosumab (N = 68) vs. Placebo (66) Woche 24
<b>Gesamtrate</b>	
<i>Jegliche TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	64 (94,1) vs. 61 (92,4)
RR <sup>a,b</sup>	1,018
[95% KI]	[0,930; 1,115]
p-Wert	0,7425
<i>Schwerwiegende TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	2 (2,9) vs. 1 (1,5)
RR <sup>a,b</sup>	1,941
[95% KI]	[0,180; 20,900]
p-Wert	1,000
<i>Schwere TEAE Grad 3 – 4</i>	
Burosumab (n (%)) vs. Placebo (n (%))	8 (11,8) vs. 8 (12,1)
RR <sup>a,b</sup>	0,971
[95% KI]	[0,387; 2,435]
p-Wert	1,000

Sicherheit	Burosumab (N = 68) vs. Placebo (66) Woche 24
<i>Therapieabbruch durch TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	0 (0) vs. 0 (0)
RR <sup>a,b</sup>	-
[95% KI]	-
p-Wert	-
<i>Tod durch TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	0 (0) vs. 0 (0)
RR <sup>a,b</sup>	-
[95% KI]	-
p-Wert	-
<b>TEAE von besonderem Interesse</b>	
<i>Reaktionen an der Einstichstelle</i>	
Burosumab (n (%)) vs. Placebo (n (%))	8 (11,8) vs. 8 (12,1)
RR <sup>a,b</sup>	0,971
[95% KI]	[0,387, 2,435]
p-Wert	1,0000
<i>Hypersensitivität</i>	
Burosumab (n (%)) vs. Placebo (n (%))	4 (5,9) vs. 4 (6,1)
RR <sup>a,b</sup>	0,971
[95% KI]	[0,253, 3,721]
p-Wert	1,0000
<i>Hyperphosphatämie</i>	
Burosumab (n (%)) vs. Placebo (n (%))	4 (5,9) vs. 0 (0)
RR <sup>a,b</sup>	8,739
[95% KI]	[0,479, 159,198]
p-Wert	0,1195
<i>Restless-Legs-Syndrom</i>	
Burosumab (n (%)) vs. Placebo (n (%))	8 (11,8) vs. 6 (9,1)
RR <sup>a,b</sup>	1,294
[95% KI]	[0,475, 3,528]
p-Wert	0,7791
<p>N: Patientenzahl der entsprechend definierten Population; n: Anzahl der Patienten mit mindestens einem Ereignis in der entsprechenden Kategorie; %: Anteil von N; TEAE: <i>Treatment emergent adverse event</i> (unerwünschtes Ereignis nach Beginn der Behandlung); [95% KI]: 95%-iges Konfidenzintervall; RR: <i>Risk Ratio</i> (relatives Risiko)</p> <p>Ein RR &lt; 1 zeigt einen Vorteil für Burosumab an und ein RR &gt; 1 einen Vorteil für Placebo.</p> <p>a: RR: Eine detaillierte Beschreibung des statistischen Modells ist in den Abschnitten 4.2.5.2 und 4.3.1.3.1.7 zu finden.</p> <p>b: Treten keine Ereignisse auf, wird ein Korrekturwert von 0,5 zu den Zellen der Vierfeldertafel hinzugefügt.</p>	

Die Studiendaten der Studie UX023-CL303 belegen die gute Verträglichkeit von Burosumab bei Erwachsenen mit XLH. Die meisten *unerwünschten Ereignisse* nach Beginn der Behandlung (*treatment emergent adverse events, TEAE*) waren mild oder moderat und traten nahezu gleichhäufig in den Behandlungsgruppen (Burosumab-Gruppe und Placebo-Gruppe) auf. Es waren keine Therapieabbrüche oder Todesfälle während der doppelblinden Phase der Studie (24 Wochen) zu verzeichnen (Tabelle 4-4).

Zu den *TEAE*, die aufgrund der pharmakologischen Eigenschaft von Burosumab zu erwarten sind, zählt die Hyperphosphatämie. Diese trat bei 4 der 68 Studienteilnehmer der Burosumab-Gruppe auf (Grad 1 und 2) und ließ sich durch eine Dosisreduktion beheben (Tabelle 4-4).

Die mit einer Phosphatsubstitution assoziierten schwerwiegenden Folgeerkrankungen wie z. B. eine Nephrokalzinose, eine Nephrolithiasis oder ein sekundärer bzw. tertiärer Hyperparathyreoidismus waren während der gesamten doppelblinden Phase der Studie unter Burosumab nicht zu beobachten.

In der Gesamtschau zeichnet sich die Burosumab-Behandlung durch ein **gutes Sicherheitsprofil** und eine **gute Verträglichkeit** aus.

### **Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen**

Bis zur Zulassung von Burosumab existierte keine Behandlung für erwachsene XLH-Patienten, die direkt in den Pathomechanismus der Erkrankung eingreift. Die einzig verfügbare Therapieoption, eine Substitutionstherapie aus oralem Phosphat und aktivem Vitamin D, wird laut der aktuellen S1-Leitlinie nur für Kinder empfohlen [3]. Darunter wird jedoch das **primäre Therapieziel**, eine **Serumphosphat-Konzentration im unteren Normbereich** (LLN  $\geq 2,5$  mg/dl) nicht erlangt. Infolgedessen treten schwerwiegende Folgeerkrankungen wie z. B. Skelettdeformationen, Arthrose, Osteomalazie, Pseudofrakturen, Enthesopathien und Spinalkanalstenosen auf. Die damit verbundenen chronischen Schmerzen und die ausgeprägte Steifheit führen zu erheblichen Einschränkungen im Alltag von XLH-Patienten. Aufgrund der unzureichenden Wirksamkeit sowie den zu beobachtenden schwerwiegenden therapieassoziierten Nebenwirkungen wird die Phosphatsubstitution häufig im Erwachsenenalter abgebrochen. Neben teils schweren Operationen erfolgt anschließend eine symptomatische Therapie mit Schmerzmitteln.

Mit Burosumab steht erstmalig eine **zielgerichtete Therapieoption** zur Verfügung, die es ermöglicht, innerhalb kurzer Zeit die **Serumphosphat-Homöostase** von Erwachsenen mit XLH **wiederherzustellen und langfristig aufrechtzuerhalten**. Die Überlegenheit von Burosumab gegenüber der zVT wird auch anhand der bisher nicht erreichten Verbesserung von weiteren patientenrelevanten Symptomen der XLH-Erkrankung deutlich. Dies beinhaltet sowohl eine **signifikante Abnahme der Steifheit und Verbesserung der physischen Funktion sowie der Gehfähigkeit**. Sicherheitsbedenken konnten anhand der Gesamtraten der unerwünschten Ereignisse nicht festgestellt werden. Zudem traten die bei einer Phosphatsubstitution häufig zu beobachtenden schwerwiegenden Nebenwirkungen (Nephrokalzinosen, Nephrolithiasis, sekundärer oder tertiärer Hyperparathyreoidismus) während der doppelblinden Phase der Studie unter Burosumab nicht auf.

Die *Orphan-Drug-Designation*, die Burosumab am 10. November 2014 durch die Europäische Arzneimittelagentur (EMA) erhalten hat, unterstreicht die fehlende therapeutische Alternative zur Behandlung von XLH-Patienten und verdeutlicht die große Versorgungslücke, die erst durch die Zulassung von Burosumab am 30.09.2020 geschlossen werden konnte [9].

In der Gesamtschau ergibt sich somit für den Wirkstoff Burosumab bei der Behandlung von Erwachsenen mit XLH, die innerhalb eines Jahres keine Verbesserung der klinischen Symptome zeigen oder bei denen eine Phosphatsubstitution kontraindiziert ist, ein **Hinweis auf einen beträchtlichen Zusatznutzen**.

## 4.2 Methodik

Abschnitt 4.2 soll die Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens beschreiben. Der Abschnitt enthält Hilfestellungen für die Darstellung der Methodik sowie einige Vorgaben, die aus den internationalen Standards der evidenzbasierten Medizin abgeleitet sind. Eine Abweichung von diesen methodischen Vorgaben ist möglich, bedarf aber einer Begründung.

### 4.2.1 Fragestellung

Nach den internationalen Standards der evidenzbasierten Medizin soll eine Bewertung unter einer definierten Fragestellung vorgenommen werden, die mindestens folgende Komponenten enthält:

- Patientenpopulation
- Intervention
- Vergleichstherapie
- Endpunkte
- Studientypen

Unter Endpunkte sind dabei alle für die frühe Nutzenbewertung relevanten Endpunkte anzugeben (d. h. nicht nur solche, die ggf. in den relevanten Studien untersucht wurden).

Die Benennung der Vergleichstherapie in Modul 4 muss zur Auswahl der zweckmäßigen Vergleichstherapie im zugehörigen Modul 3 konsistent sein.

*Geben Sie die Fragestellung der vorliegenden Aufarbeitung von Unterlagen zur Untersuchung des medizinischen Nutzens und des medizinischen Zusatznutzens des zu bewertenden Arzneimittels an. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.*

Das vorliegende Dossier bezieht sich auf den Nachweis der Wirksamkeit und Sicherheit des Wirkstoffs Burosumab (Crysvita<sup>®</sup>), welches zur Behandlung von erwachsenen Patienten mit X-chromosomaler Hypophosphatämie (XLH) angezeigt ist. Es handelt sich hierbei um ein Arzneimittel zur Behandlung eines seltenen Leidens gemäß § 35a Abs. 1 Satz 11 SGB V. Der Zusatznutzen von Burosumab (Crysvita<sup>®</sup>) ist somit bereits durch die Zulassung belegt. Nachweise über den medizinischen Nutzen und Zusatznutzen im Verhältnis zu einer zuvor bestimmten zweckmäßigen Vergleichstherapie (zVT) müssen demnach nicht erbracht werden. Lediglich das Ausmaß des Zusatznutzen ist für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, zu beschreiben.

Übersteigt der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten zwölf Kalendermonaten

einen Betrag von 50 Millionen Euro, hat der pharmazeutische Unternehmer gemäß § 35a Absatz 1 Satz 12 SGB V innerhalb von drei Monaten nach Aufforderung durch den G-BA Nachweise gemäß 5. Kapitel § 5 Absatz 1 bis 6 VerfO des G-BA zu übermitteln.

Der G-BA hat mit Schreiben vom 21.10.2021 die Kyowa Kirin GmbH darüber informiert, dass der Umsatz mit Burosumab (Crysvita®) zu Lasten der GKV einen Betrag von 50 Millionen Euro überschritten hat. Maßgeblicher Zeitpunkt für die Einreichung der Nachweise gemäß § 35a Abs. 1 Satz 12 SGB V ist damit der 01.02.2022. Im vorliegenden Dossier werden nun die erforderlichen Nachweise des Zusatznutzens von Burosumab gegenüber der zVT „best supportive care“ dargestellt.

Die im vorliegenden Dossier zu beantwortende Frage lautet:

Wie groß ist das Ausmaß des Zusatznutzens von Burosumab (Crysvita®) bezüglich patientenrelevanter Endpunkte bei der Behandlung von Erwachsenen mit XLH, die unter einer Phosphatsubstitution innerhalb eines Jahres keine Verbesserung der klinischen Symptome zeigen oder bei denen eine Phosphatsubstitution kontraindiziert ist?

Es ist zu beachten, dass bei Erwachsenen mit XLH zwischen zwei Teilpopulationen unterschieden wird, die sich hinsichtlich des Ansprechens auf eine Substitutionstherapie aus oralem Phosphat und aktivem Vitamin D unterscheiden:

### 1. Teilpopulation

Erwachsene, die innerhalb **eines Jahres** auf eine Phosphatsubstitution **ansprechen**

### 2. Teilpopulation

Erwachsene, die innerhalb **eines Jahres nicht** auf eine Phosphatsubstitution **ansprechen**

Trotz einer langjährigen Phosphatsubstitution wird bei nahezu allen Erwachsenen das primäre Therapieziel, eine Serumphosphat-Konzentration im unteren Normbereich ( $\geq 2,5$  mg/dl), nicht erlangt. Zudem lässt sich bei diesen XLH-Patienten das Fortschreiten der Erkrankung nicht aufhalten. Unterstützt wird dies auch durch die Charakteristika der Studienteilnehmer der Studie UX023-CL303 [5]. Es kann somit davon ausgegangen werden, dass nahezu alle behandlungsbedürftigen Erwachsenen mit XLH der Teilpopulation zuzuordnen sind, die nicht auf eine Phosphatgabe ansprechen.

## **Patientenpopulation**

Die Patientenpopulation umfasst Erwachsene mit XLH ( $\geq 18$  Jahre), die unter einer Phosphatsubstitution innerhalb von einem Jahr keine Verbesserung der klinischen Symptome zeigen. Zudem sind auch Erwachsene mit XLH eingeschlossen, bei denen eine Phosphatsubstitution kontraindiziert ist. Eine ausführliche Begründung ist in Abschnitt 4.1 und Modul 3 Abschnitt 3.2 enthalten.

## Intervention

Die für die Nutzenbewertung relevante Intervention ist Burosumab, ein rekombinanter, vollständig humaner monoklonaler Antikörper (IgG1), der spezifisch den phosphaturischen Fibroblasten-Wachstumsfaktor 23 (FGF23) bindet und dadurch dessen biologische Aktivität hemmt. Durch das direkte Eingreifen in den molekularen Pathomechanismus, der der XLH-Erkrankung zugrunde liegt, stellt Burosumab den ersten Wirkstoff dar, der dauerhaft die Serumphosphat-Homöostase wiederherstellt.

## Zweckmäßige Vergleichstherapie

Die zweckmäßige Vergleichstherapie (zVT) für das Anwendungsgebiet XLH wurde in einem Beratungsgespräch am 11.03.2020 beim G-BA erfragt. Der G-BA gab in diesem Gespräch als zVT eine Phosphatsubstitution mit aktivem Vitamin D an [7]. Die Kyowa Kirin GmbH folgt diesen Angaben jedoch nicht, da nach eigener Auffassung bis zum heutigen Zeitpunkt keine Leitlinien oder einheitliche Therapieempfehlungen zur Behandlung von Erwachsenen mit XLH vorliegen. Weiterhin weisen, trotz mehrjähriger Vorbehandlung mit Phosphat, die meisten Studienteilnehmer der Studie UX023-CL303 schwere Folgeerkrankungen auf wie z. B. Skelettdeformationen, Arthrose, Osteomalazie, Pseudofrakturen und Enthesopathien. Aufgrund der mangelnden Wirksamkeit stellt die Phosphatsubstitution eine unzureichende Therapie dar, deren Fortführung nicht einer adäquaten Umsetzung der zVT entsprechen kann. Da keine den Pathomechanismus adressierende Therapieoption existiert, ist die Kyowa Kirin GmbH der Meinung, dass lediglich „*best supportive care*“ die alleinige zweckmäßige Vergleichstherapie darstellt. Als „*best supportive care*“ wird die Therapie verstanden, die eine bestmögliche, patientenindividuell optimierte unterstützende Behandlung zur Linderung von Symptomen und Verbesserung der Lebensqualität gewährleistet. Im Anwendungsgebiet Erwachsene mit XLH wird Placebo in Kombination mit einer Begleittherapie als geeignete Umsetzung von „*best supportive care*“ angesehen. Eine ausführliche Begründung ist in Modul 3 Abschnitt 3.1.2 zu finden.

## Patientenrelevante Endpunkte

Zur Beurteilung des Ausmaßes des Zusatznutzens wurden die patientenrelevanten Endpunkte im Anwendungsgebiet von XLH betrachtet. Diese umfassen sowohl die Morbidität als auch die Sicherheit. Da es sich bei XLH um eine schwerwiegende chronische Erkrankung handelt, die aber potenziell nicht tödlich verläuft, wird in der Studie UX023-CL303 die Mortalität durch die Datenerhebung der schweren unerwünschten Ereignisse erhoben und deshalb nicht gesondert dargestellt.

## Studientyp

Für die Nutzenbewertung von Arzneimitteln sollen ausschließlich Studien mit hoher Evidenz und niedrigem Verzerrungspotenzial eingeschlossen werden. Deshalb werden im Folgenden lediglich Studien berücksichtigt, deren Studiendesign diesen Kriterien entspricht. Hierzu zählen randomisierte kontrollierte Studien, die in dem Anwendungsgebiet XLH mit einer Mindestlaufzeit von 24 Wochen durchgeführt wurden.

#### 4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung

Die Untersuchung der in Abschnitt 4.2.1 benannten Fragestellung soll auf Basis von klinischen Studien vorgenommen werden. Für die systematische Auswahl von Studien für diese Untersuchung sollen Ein- und Ausschlusskriterien für die Studien definiert werden. Dabei ist zu beachten, dass eine Studie nicht allein deshalb ausgeschlossen werden soll, weil keine in einer Fachzeitschrift veröffentlichte Vollpublikation vorliegt. Eine Bewertung der Studie kann beispielsweise auch auf Basis eines ausführlichen Ergebnisberichts aus einem Studienregister/ einer Studienergebnisdatenbank erfolgen, während ein Kongressabstrakt allein in der Regel nicht für eine Studienbewertung ausreicht.

*Benennen Sie die Ein- und Ausschlusskriterien für Studien zum medizinischen Nutzen und Zusatznutzen. Machen Sie dabei mindestens Aussagen zur Patientenpopulation, zur Intervention, zur Vergleichstherapie, zu den Endpunkten, zum Studientyp und zur Studiendauer und begründen Sie diese. Stellen Sie die Ein- und Ausschlusskriterien zusammenfassend in einer tabellarischen Übersicht dar. Erstellen Sie dabei für unterschiedliche Themen der Recherche (z. B. unterschiedliche Fragestellungen) jeweils eine separate Übersicht.*

Die Bewertung des Ausmaßes des Zusatznutzens von Burosumab (Crysvita®) erfolgt auf Grundlage der Zulassungsstudie UX023-CL303 (NCT02526160 [10]; EudraCT 2014-005529-11 [11]; WHO ICTRP 2014-005529-11 [12]).

Um die vorhandene Evidenz von Burosumab in dem vorliegenden Anwendungsgebiet vollständig abzubilden, wird zusätzlich eine systematische Literaturrecherche in Literaturdatenbanken und Studienregistern durchgeführt.

Die Ein- und Ausschlusskriterien der Studien, die für die Bewertung des Ausmaßes des Zusatznutzens von Burosumab herangezogen werden, orientieren sich an der zuvor definierten Fragestellung gemäß Abschnitt 4.2.1. Im vorliegenden Dossier soll sowohl der medizinische Nutzen als auch der Zusatznutzen primär anhand direkter Evidenz nachgewiesen werden. Bei den Ein- und Ausschlusskriterien müssen folgende Parameter (gemäß PICOS = *Population, Intervention, Comparison, Outcome, Study design*) berücksichtigt werden die anschließend detailliert beschrieben werden:

- Patientenpopulation
- Intervention
- Vergleichstherapie (Komparator)
- Endpunkte
- Studientyp/-dauer

## Ein- und Ausschlusskriterien für randomisierte kontrollierte Studien mit dem zu bewertenden Arzneimittel

Im Folgenden sind die Ein- und Ausschlusskriterien für randomisierte und kontrollierte Studien mit Burosumab als Intervention aufgeführt (Tabelle 4-5 und Tabelle 4-6).

Tabelle 4-5: Einschlusskriterien für RCT mit Burosumab als Intervention

Einschlusskriterien		
<b>E1</b>	<b>Patientenpopulation</b>	Erwachsene Patienten ( $\geq 18$ Jahre) mit X-chromosomaler Hypophosphatämie (XLH) <sup>a</sup>
<b>E2</b>	<b>Intervention</b>	Burosumab, Crysvita®, KRN23 1,0 mg/kg KG alle 4 Wochen Höchstdosis: 90 mg Subkutane Applikation
<b>E3</b>	<b>Komparator</b>	Placebo + Begleittherapie (= <i>best supportive care</i> ) <sup>b</sup>
<b>E4</b>	<b>Endpunkte</b>	Erhebung von mindestens einem patientenrelevanten Endpunkt, wie zum Beispiel Morbidität (z. B. Schmerz, Fatigue, Steifheit), Lebensqualität und unerwünschte Ereignisse
<b>E5</b>	<b>Studiendesign</b>	Randomisierte, kontrollierte Studien (RCT) Doppelblind
<b>E6</b>	<b>Studiendauer</b>	$\geq 24$ Wochen
<b>E7</b>	<b>Publikationstyp</b>	Volltextpublikation oder Studienbericht Veröffentlichung in deutscher oder englischer Sprache
<p>a: Anhand der Literatur und den Charakteristika der Studienteilnehmer der Studie UX023-CL303 kann davon ausgegangen werden, dass nahezu alle erwachsenen XLH-Patienten der Teilpopulation <i>Erwachsene, die innerhalb eines Jahres nicht auf eine Phosphatsubstitution ansprechen</i> zu zuordnen sind. Eine Einschränkung bei der Suche nach Studien im Anwendungsgebiet ist daher nicht notwendig.</p> <p>b: Der G-BA hat für das vorliegende Anwendungsgebiet als zVT eine Phosphatsubstitution bestimmt [7]. Nach Auffassung der Kyowa Kirin GmbH stellt die Phosphatsubstitution für Erwachsene mit XLH jedoch eine unzureichende Therapie dar. Da keine weiteren Therapieoptionen zur Verfügung stehen, kommt für erwachsene XLH-Patienten daher lediglich „<i>best supportive care</i>“ als zVT in Betracht. Als geeignete Umsetzung von „<i>best supportive care</i>“ wird Placebo in Kombination mit Begleittherapien gesehen. Eine ausführliche Begründung ist in Modul 3 Abschnitt 3.1.2 aufgeführt. Um die formale Vollständigkeit zu gewährleisten, wurde dennoch eine Suche nach Studien mit dem Komparator Phosphat durchgeführt. Es konnten jedoch keine Studien identifiziert werden, die diesen historischen Therapieansatz verwenden und gleichzeitig die in der Tabelle aufgeführten Einschlusskriterien (E1 – E2) und (E4 – E7) erfüllen.</p>		

Tabelle 4-6: Ausschlusskriterien für RCT mit Burosumab als Intervention

Ausschlusskriterien		
A1	Patientenpopulation nicht entsprechend E1	Patienten mit XLH < 18 Jahren
A2	Intervention nicht entsprechend E2	Andere Wirkstoffe Dosierung, Applikation nicht gemäß Zulassung von Burosumab
A3	Komparator	Andere Wirkstoffe
A4	Endpunkte nicht entsprechend E4	Studien, bei denen ausschließlich Surrogat-Parameter evaluiert werden
A5	Studiendesign nicht entsprechend E5	Nicht-randomisierte, kontrollierte Studien
A6	Studiendauer nicht entsprechend E6	< 24 Wochen
A7	Publikationstyp nicht entsprechend E7	Keine Primärpublikation, Review, Abstract, Kongress-Poster Andere Sprache als Deutsch oder Englisch

### Erläuterungen zu den Ein- und Ausschlusskriterien

#### *Patientenpopulation (Ein- und Ausschlusskriterium 1)*

Für die Bewertung des Zusatznutzens sollten lediglich randomisierte kontrollierte Studien berücksichtigt werden, die erwachsene Patienten ( $\geq 18$  Jahre) mit dem Krankheitsbefund XLH einschließen. Dies beinhaltet Erwachsene, bei denen die XLH-Erkrankung entweder durch eine Mutationsanalyse des *PHEX*-Gens (*Phosphat regulierendes Gen mit Homologie zu Endopeptidasen auf dem X-Chromosom*) oder anhand einer erhöhten Serumkonzentration des Fibroblasten-Wachstumsfaktors 23 (FGF23) diagnostiziert wurde und charakteristische Merkmale der Erkrankung wie z. B. Skelettdeformationen sowie Knochen- und Gelenkschmerzen aufweisen.

In der Nutzenbewertung werden nur Studien mit erwachsenen XLH-Patienten im Alter  $\geq 18$  Jahren eingeschlossen, die der im Abschnitt 4.2.1 definierten Zielpopulation entsprechen.

#### *Intervention (Ein- und Ausschlusskriterium 2)*

Für die Bewertung des Ausmaßes des Zusatznutzens von Burosumab bei XLH werden nur Studien herangezogen, in denen Burosumab als Intervention angewendet wird.

Die empfohlene Anfangsdosis von Burosumab beträgt 1,0 mg/kg Körpergewicht (KG). Burosumab wird alle vier Wochen subkutan verabreicht. Die Maximaldosis je Injektion beträgt 90 mg. Hierbei ist zu beachten, dass die zu applizierenden Dosen auf die nächsten 10 mg auf- bzw. abgerundet werden.

#### *Komparator (Ein- und Ausschlusskriterium 3)*

Literaturrecherchen ergaben, dass im vorliegenden Anwendungsgebiet bis zum heutigen Tag keine Leitlinien oder einheitlichen Therapieempfehlungen für Erwachsene mit XLH vorliegen. Die einzige S1-Leitlinie „Hereditäre hypophosphatämische Rachitiden“ der Deutschen

Gesellschaft für Kinderendokrinologie und -diabetologie (DGKED) e.V. [3] bezieht sich lediglich auf die Behandlung von Kindern und Jugendlichen und sieht eine Phosphatsubstitution zusammen mit aktivem Vitamin D vor. Trotz einer optimal eingestellten Phosphatgabe wird der untere Normbereich der Serumphosphat-Konzentration für Erwachsene nicht erreicht, sodass das Auftreten von schwerwiegenden Folgeerkrankungen der XLH sowie therapieassoziierten Sekundärerkrankungen nicht vermieden werden kann. Aufgrund der starken Nebenwirkungen und dem unzureichenden Ansprechen auf die Therapie wird diese häufig nach Abschluss des Skelettwachstums abgebrochen. Anschließend steht für die XLH-Patienten lediglich eine symptomatische Therapie zur Verfügung. Dies beinhaltet sowohl den Einsatz von Schmerzmitteln als auch schwere Operationen, um die frühzeitigen Verschleißerscheinungen, die Brüche und die neu auftretenden Verformungen sowie die in der Kindheit erworbenen Fehlstellungen zu beheben. Da weder Leitlinien noch einheitliche Therapieempfehlungen für die Behandlung von Erwachsenen mit XLH vorliegen, stellt deshalb lediglich „*best supportive care*“ einen angemessenen Komparator da. Als „*best supportive care*“ wird die Therapie verstanden, die eine bestmögliche, patientenindividuell optimierte unterstützende Behandlung zur Linderung von Symptomen und Verbesserung der Lebensqualität gewährleistet. Im Anwendungsgebiet Erwachsene mit XLH wird Placebo in Kombination mit einer Begleittherapie, wie in der Studie UX023-CL303 eingesetzt, als geeignete Umsetzung von „*best supportive care*“ angesehen.

#### ***Endpunkte (Ein- und Ausschlusskriterium 4)***

Um in der Indikation XLH bei Erwachsenen das Ausmaß des Zusatznutzens von Burosumab zu bestimmen, werden krankheitsspezifische patientenrelevante Endpunkte analysiert. Diese beinhalten sowohl die Morbidität als auch die Sicherheit. Eine Übersicht über die patientenrelevanten Endpunkte bei Erwachsenen mit XLH ist in Abschnitt 4.2.5.2 dargestellt. Die Relevanz und Validität der entsprechenden Endpunkte werden in Abschnitt 4.2.5.2 detailliert begründet. Da es sich bei XLH um eine schwerwiegende chronische Erkrankung handelt, die aber nicht potenziell tödlich verläuft, wird in der Studie UX023-CL303 die Mortalität durch die Datenerhebung der schweren unerwünschten Ereignisse erhoben und deshalb nicht gesondert dargestellt.

#### ***Studientyp (Ein-/Ausschlusskriterium 5)***

Für die Nutzenbewertung von Burosumab werden ausschließlich RCTs herangezogen, die einen direkten Vergleich mit der zVT „*best supportive care*“ (Placebo + Begleittherapie) ermöglichen.

#### ***Studiendauer (Ein-/Ausschlusskriterium 6)***

In früheren Nutzenbewertungsverfahren wurde, um das Ansprechen bei einer chronischen Erkrankung zu evaluieren, eine Mindestlaufzeit der Studie von 24 Wochen vorausgesetzt. Relevante Verfahren stellen u. a. die Nutzenbewertung von Secukinumab bei Plaque-Psoriasis oder Ivacaftor bei zystischer Fibrose dar („Insgesamt sind für die Nutzenbewertung im Anwendungsgebiet der zystischen Fibrose Studien von mindestens 24 Wochen notwendig für die Gegenüberstellung von Nutzen und Schaden“) [13, 14]. Darüber hinaus zeigten Fallberichte von Patienten mit XLH oder ernährungsbedingter Osteomalazie, dass eine Heilung der

Osteomalazie sowie Verbesserung der Krankheits-assoziierten Symptome wie z. B. Schmerz 12 – 24 Wochen nach Wiederherstellung der Serumphosphat-Homöostase eintritt [15, 16]. Aufgrund dessen wird eine Laufzeit der Studie von 24 Wochen als ausreichend angesehen, um die Wirksamkeit von Burosumab auf die Erkrankung XLH zu untersuchen.

#### ***Publikationstyp (Ein-/ Ausschlusskriterium 7)***

Es werden ausschließlich Studien berücksichtigt, für die eine Vollpublikation oder ein Studienbericht in deutscher oder englischer Sprache vorliegen, die die Kriterien des CONSORT-Statements erfüllen.

### **4.2.3 Informationsbeschaffung**

In den nachfolgenden Abschnitten ist zu beschreiben, nach welcher Methodik Studien identifiziert wurden, die für die Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens in dem in diesem Dokument bewerteten Anwendungsgebiet herangezogen werden. Dies bezieht sich sowohl auf publizierte als auch auf unpublizierte Studien. Die Methodik muss dazu geeignet sein, die relevanten Studien (gemäß den in Abschnitt 4.2.2 genannten Kriterien) systematisch zu identifizieren (systematische Literaturrecherche).

#### **4.2.3.1 Studien des pharmazeutischen Unternehmers**

Für die Identifikation der Studien des pharmazeutischen Unternehmers ist keine gesonderte Beschreibung der Methodik der Informationsbeschaffung erforderlich. Die vollständige Auflistung aller Studien, die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie aller Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, erfolgt in den Abschnitten 4.3.1 und 4.3.2, jeweils im Unterabschnitt „Studien des pharmazeutischen Unternehmers“. Die Darstellung soll auf Studien mit Patienten in dem Anwendungsgebiet, für das das vorliegende Dokument erstellt wird, beschränkt werden.

#### **4.2.3.2 Bibliografische Literaturrecherche**

Die Durchführung einer bibliografischen Literaturrecherche ist erforderlich, um sicherzustellen, dass ein vollständiger Studienpool in die Bewertung einfließt.

Eine bibliografische Literaturrecherche muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine bibliografische Literaturrecherche immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird.

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die bibliografische Literaturrecherche soll mindestens in den Datenbanken MEDLINE (inklusive „in-process & other non-indexed citations“) und EMBASE sowie in der Cochrane-

Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Datenbanken (z. B. CINAHL, PsycINFO etc.) durchgeführt werden.

Die Suche soll in jeder Datenbank einzeln und mit einer für die jeweilige Datenbank adaptierten Suchstrategie durchgeführt werden. Die Suchstrategien sollen jeweils in Blöcken, insbesondere getrennt nach Indikation, Intervention und ggf. Studientypen, aufgebaut werden. Wird eine Einschränkung der Strategien auf bestimmte Studientypen vorgenommen (z. B. randomisierte kontrollierte Studien), sollen aktuelle validierte Filter hierfür verwendet werden. Alle Suchstrategien sind in Anhang 4-A zu dokumentieren.

*Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Datenbanken eine bibliografische Literaturrecherche durchgeführt wurde. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, wenn bei der Recherche generelle Einschränkungen vorgenommen wurden (z. B. Sprach- oder Jahreseinschränkungen), und begründen Sie diese.*

Die bibliografische Literaturrecherche wurde in den Datenbanken MEDLINE und EMBASE sowie in der Cochrane-Datenbank durchgeführt. Für MEDLINE (inklusive *in-process, in-data-review & other non-indexed citations*) und EMBASE erfolgte die Suche über die Oberfläche von Ovid, für die Cochrane-Datenbank wurde die Cochrane-eigene Suchoberfläche *Cochrane Central Register of Controlled Trials (Clinical Trails)* verwendet. Die Suche beschränkte sich auf publizierte Ergebnisse aus RCTs mit dem zu bewertenden Arzneimittel und erfolgte am 02.11.2021.

Die verwendeten Suchstrategien und Suchbegriffe sind in Anhang 4-A detailliert beschrieben und dokumentiert. Für die Suche wurden Synonyme des zu bewertenden Arzneimittels Burosumab in Kombination von Freitext und Begriffen der *Medical Subject Headings (MeSH)* verwendet. Eine Jahreseinschränkung wurde nicht vorgenommen. Für MEDLINE und EMBASE wurde zur spezifischen Suche nach RCTs der validierte WONG-Filter verwendet [17].

#### **4.2.3.3 Suche in Studienregistern/ Studienergebnisdatenbanken**

Eine Suche in öffentlich zugänglichen Studienregistern/ Studienergebnisdatenbanken ist grundsätzlich durchzuführen, um sicherzustellen, dass laufende Studien sowie abgeschlossene Studien auch von Dritten vollständig identifiziert werden und in Studienregistern / Studienergebnisdatenbanken vorliegende Informationen zu Studienmethodik und –ergebnissen in die Bewertung einfließen.

Eine Suche in Studienregistern/ Studienergebnisdatenbanken muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine Suche in Studienregistern sowie Studienergebnisdatenbanken immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird.

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Suche soll mindestens in den Studienregistern/ Studienergebnisdatenbanken [clinicaltrials.gov](http://www.clinicaltrials.gov) ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)), EU Clinical Trials Register (EU-CTR, [www.clinicaltrialsregister.eu](http://www.clinicaltrialsregister.eu)), International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal), Suchportal der WHO, Clinical Data Suchportal der European Medicines Agency (<https://clinicaldata.ema.europa.eu>) sowie dem Arzneimittel-Informationssystem (AMIS, <https://www.pharmnet-bund.de/dynamic/de/arsneimittel-informationssystem/index.html>) durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Studienregistern / Studienergebnisdatenbanken (z. B. krankheitsspezifische Studienregister oder Studienregister einzelner pharmazeutischer Unternehmen) durchgeführt werden. Die Suche in Studienregistern/ Studienergebnisdatenbanken anderer pharmazeutischer Unternehmer ist insbesondere bei indirekten Vergleichen sinnvoll, wenn Studien zu anderen Arzneimitteln identifiziert werden müssen.

Die Suche soll in jedem Studienregister/ Studienergebnisdatenbank einzeln und mit einer für das jeweilige Studienregister/ Studienergebnisdatenbank adaptierten Suchstrategie durchgeführt werden. Die Suche soll abgeschlossene, abgebrochene und laufende Studien erfassen. Alle Suchstrategien sind in Anhang 4-B zu dokumentieren.

Für Clinical Data (Suchportal der European Medicines Agency) und das Arzneimittel-Informationssystem (AMIS) genügt hingegen die Suche nach Einträgen mit Ergebnisberichten zu Studien, die bereits anderweitig (z.B. über die bibliografische Literaturrecherche und Studienregistersuche) identifiziert wurden. Eine Dokumentation der zugehörigen Suchstrategie ist nicht erforderlich.

*Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Studienregistern/ Studienergebnisdatenbanken die Suche durchgeführt wurde. Begründen Sie dabei Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, wenn bei der Recherche generelle Einschränkungen vorgenommen wurden (z. B. Jahreseinschränkungen), und begründen Sie diese.*

Zusätzlich zu der systematischen Literaturrecherche wurde eine weitere Recherche in öffentlich zugänglichen Studienregistern durchgeführt. Diese hat zum Ziel, alle klinischen Studien, die sich für die Bewertung des Ausmaßes des Zusatznutzens von Burosumab in der Indikation Erwachsene mit XLH eignen, zu identifizieren und beinhaltet sowohl alle abgeschlossenen als auch abgebrochenen sowie laufende Studien.

Die Suche wurde in den oben aufgeführten Studienregistern [clinicaltrials.gov](http://www.clinicaltrials.gov) (<https://www.clinicaltrials.gov/>), *WHO International Clinical Trials Registry Plattform* (WHO ICTRP, <https://trialsearch.who.int/Default.aspx>), *EU Clinical Trials Registry* (EU-CTR, <https://www.clinicaltrialsregister.eu/>) durchgeführt. Die detaillierte Suchstrategie ist für jedes Studienregister im Anhang 4-B getrennt dargestellt. Es wurden sowohl der Wirkstoffname als

auch entsprechende Synonyme verwendet. Die Suchstrategie wurde für jedes Studienregister angepasst.

Zusätzlich wurde im Clinical Data Suchportal der EMA (<https://clinicaldata.ema.europa.eu/web/cdp/home>) sowie im Arzneimittel-Informationssystem AMIce (ehemals AMIS) nach ergänzenden Einträgen und Ergebnisberichten gesucht. Für die Suche im *Clinical Data* Suchportal der EMA wurde als Suchbegriff das Schlagwort „UX023-CL303“ verwendet, für die Suche im Arzneimittel-Informationssystem AMIce hingegen der Wirkstoffname sowie die entsprechenden Synonyme.

Die Suche wurde am 02.11.2021 durchgeführt; es wurden keine generellen Einschränkungen bei den genannten Recherchen vorgenommen.

#### 4.2.3.4 Suche auf der Internetseite des G-BA

Die Internetseite des G-BA ist grundsätzlich zu durchsuchen, um sicherzustellen, dass alle vorliegenden Daten zu Studienmethodik und –ergebnissen von relevanten Studien in die Bewertung einfließen.

Auf der Internetseite des G-BA werden Dokumente zur frühen Nutzenbewertung nach §35a SGB V veröffentlicht. Diese enthalten teilweise anderweitig nicht veröffentlichte Daten zu Studienmethodik und –ergebnissen<sup>1</sup>. Solche Daten sind dabei insbesondere in den Modulen 4 der Dossiers pharmazeutischer Unternehmer, in IQWiG-Nutzenbewertungen sowie dem Beschluss des G-BA einschließlich der Tragenden Gründe und der Zusammenfassenden Dokumentation zu erwarten.

Die Suche auf der Internetseite des G-BA muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine Suche auf der G-BA Internetseite immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird. Die Suche ist dann sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie durchzuführen. Es genügt die Suche nach Einträgen zu Studien, die bereits anderweitig (z.B. über die bibliografische Literaturrecherche und Studienregistersuche) identifiziert wurden. Eine Dokumentation der zugehörigen Suchstrategie ist nicht erforderlich.

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

---

<sup>1</sup> Köhler M, Haag S, Biester K, Brockhaus AC, McGauran N, Grouven U, Kölsch H, Seay U, Hörn H, Moritz G, Staack K, Wieseler B. Information on new drugs at market entry: retrospective analysis of health technology assessment reports, journal publications, and registry reports. *BMJ* 2015;350:h796

*Beschreiben Sie nachfolgend das Vorgehen für die Suche. Benennen Sie die Wirkstoffe und die auf der Internetseite des G-BA genannten zugehörigen Vorgangsnummern, zu denen Sie eine Suche durchgeführt haben.*

*Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.*

Des Weiteren erfolgte eine Recherche auf der Internetseite des G-BA, um weitere Studien, die den Kriterien in Abschnitt 4.2.2 entsprechen, zu identifizieren. Dafür wurde über die Suchmaske der Website des G-BA mit dem Schlagwort „Burosumab“ gesucht. Zusätzlich wurde nach dem Schlagwort „UX023-CL303“ gesucht, um eventuell weitere Einträge zu der zuvor identifizierten Studie UX023-CL303 zu identifizieren. Der Wirkstoff Burosumab wurde bereits vom G-BA bewertet (Erstbewertung: 2020-11-01-D-588). Die Suche wurde am 09.11.2021 durchgeführt.

#### **4.2.3.5 Selektion relevanter Studien**

*Beschreiben Sie das Vorgehen bei der Selektion relevanter Studien aus dem Ergebnis der in den Abschnitten 4.2.3.2, 4.2.3.3 und 4.2.3.4 beschriebenen Rechenschritte. Begründen Sie das Vorgehen, falls die Selektion nicht von zwei Personen unabhängig voneinander durchgeführt wurde.*

Die Literaturrecherche wurde von zwei Personen unabhängig voneinander durchgeführt. Es wurden die Titel und Abstracts der gefundenen Treffer gescreent, die Dubletten entfernt und die Studien anhand der in Abschnitt 4.2.2 dargelegten und vordefinierten Kriterien selektiert.

Die Ergebnisse der Studienselektion sind im Abschnitt 4.3.1.1 aufgeführt.

#### **4.2.4 Bewertung der Aussagekraft der Nachweise**

Zur Bewertung der Aussagekraft der im Dossier vorgelegten Nachweise sollen Verzerrungsaspekte der Ergebnisse für jede eingeschlossene Studie beschrieben werden, und zwar separat für jeden patientenrelevanten Endpunkt. Dazu sollen insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Aspekte systematisch extrahiert werden (zur weiteren Erläuterung der einzelnen Aspekte siehe Bewertungsbogen in Anhang 4-F):

##### A: Verzerrungsaspekte der Ergebnisse auf Studienebene

- Erzeugung der Randomisierungssequenz (*bei randomisierten Studien*)
- Verdeckung der Gruppenzuteilung (*bei randomisierten Studien*)
- zeitliche Parallelität der Gruppen (*bei nicht randomisierten vergleichenden Studien*)
- Vergleichbarkeit der Gruppen bzw. Berücksichtigung prognostisch relevanter Faktoren (*bei nicht randomisierten vergleichenden Studien*)
- Verblindung des Patienten sowie der behandelnden Personen
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene

- Verblindung der Endpunkterheber
- Umsetzung des ITT-Prinzips
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

Für randomisierte Studien soll darüber hinaus das Verzerrungspotenzial bewertet und als „niedrig“ oder „hoch“ eingestuft werden. Ein niedriges Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

Eine zusammenfassende Bewertung der Verzerrungsaspekte soll nicht für nicht randomisierte Studien erfolgen.

Für die Bewertung eines Endpunkts soll für randomisierte Studien zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter A aufgeführten Aspekte als „niedrig“ oder „hoch“ eingestuft werden. Falls diese Einstufung als „hoch“ erfolgt, soll das Verzerrungspotenzial für den Endpunkt in der Regel auch als „hoch“ bewertet werden, Abweichungen hiervon sind zu begründen. Ansonsten sollen die unter B genannten endpunktspezifischen Aspekte Berücksichtigung finden.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ soll nicht zum Ausschluss der Daten führen. Die Klassifizierung soll vielmehr der Diskussion heterogener Studienergebnisse und der Einschätzung der Aussagekraft der Nachweise dienen. Für nicht randomisierte Studien können für solche Diskussionen einzelne Verzerrungsaspekte herangezogen werden.

*Beschreiben Sie die für die Bewertung der Verzerrungsaspekte und des Verzerrungspotenzials eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.*

Für das Krankheitsbild X-chromosomale Hypophosphatämie (XLH) bei Erwachsenen liegt eine abgeschlossene, randomisierte, doppelblinde, Placebo-kontrollierte und multizentrische Phase III-Studie mit dem zu bewertenden Arzneimittel Burosumab vor. Dementsprechend wird in dem vorliegenden Dossier eine zusammenfassende Bewertung der Verzerrungsaspekte auf Studien- und Endpunktebene dargestellt.

Zur Bewertung der Aussagekraft wird das endpunktübergreifende Verzerrungspotenzial zunächst auf Studienebene als hoch oder niedrig eingestuft. Im Anschluss daran wird jeder Endpunkt einzeln bewertet.

### **Verzerrungspotenzial auf Studienebene**

Um das Verzerrungspotenzial auf Studienebene als „niedrig“ oder „hoch“ einzustufen, wurden die Angaben des Studienprotokolls und des Studienberichts herangezogen und folgende Aspekte beurteilt:

- Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Verblindung des Patienten sowie der behandelnden Personen
- Ergebnisunabhängige Berichterstattung
- Sonstige Aspekte auf Studienebene

### **Verzerrungspotenzial auf Endpunktebene**

Um das Verzerrungspotenzial auf Endpunktebene als „niedrig“ oder „hoch“ einzustufen, wurden folgende Aspekte beurteilt:

- Verblindung der Endpunkterheber
- Adäquate Umsetzung des ITT-Prinzips
- Ergebnisunabhängige Berichterstattung
- Sonstige Aspekte auf Endpunktebene

Die Ergebnisse der Analysen der RCT sind in Anhang 4-F dokumentiert.

## **4.2.5 Informationssynthese und -analyse**

### **4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien**

Das Design und die Methodik der eingeschlossenen Studien soll in den Abschnitten 4.3.1 und 4.3.2, jeweils in den Unterabschnitten „Charakteristika der in die Bewertung eingeschlossenen Studien“ und den dazugehörigen Anhängen, dargestellt werden. Die Darstellung der Studien soll für randomisierte kontrollierte Studien mindestens die Anforderungen des CONSORT-Statements erfüllen (Items 2b bis 14, Informationen aus dem CONSORT-Flow-Chart)<sup>2</sup>. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-<sup>3</sup> bzw. STROBE-

---

<sup>2</sup> Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.

<sup>3</sup> Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Publ Health 2004; 94(3): 361-366.

Statements<sup>4</sup> folgen. Design und Methodik weiterer Untersuchungen sollen gemäß den verfügbaren Standards dargestellt werden.

*Beschreiben Sie, nach welchen Standards und mit welchen Informationen (Items) Sie das Design und die Methodik der eingeschlossenen Studien in Modul 4 dargestellt haben. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.*

Das Design und die Methodik der Studie, die in dieser Nutzenbewertung zum Ableiten des Zusatznutzens herangezogen wird, ist in Abschnitt 4.2.5.1 und in Abschnitt 4.3.1 detailliert beschrieben. Dabei handelt es sich um die Studie UX023-CL303 mit der Evidenzstufe 1b. Dementsprechend erfolgt die Darstellung dieser Studie anhand der *Consolidated Standards of Reporting Trials (CONSORT)*-Statements (Items 2b – 14). Zur Veranschaulichung des Patientenflusses wird zusätzlich das entsprechende CONSORT-Flow-Chart dargestellt (Abbildung 9). Nicht randomisierte Interventionsstudien und epidemiologische Beobachtungsstudien werden dagegen nicht berücksichtigt.

#### **4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien**

Die Ergebnisse der einzelnen Studien sollen in den Abschnitten 4.3.1 und 4.3.2 in den entsprechenden Unterabschnitten zunächst für jede eingeschlossene Studie separat dargestellt werden. Die Darstellung soll die Charakteristika der Studienpopulationen sowie die Ergebnisse zu allen in den eingeschlossenen Studien berichteten patientenrelevanten Endpunkten (Verbesserung des Gesundheitszustands, Verkürzung der Krankheitsdauer, Verlängerung des Überlebens, Verringerung von Nebenwirkungen, Verbesserung der Lebensqualität) umfassen. Anforderungen an die Darstellung werden in den Unterabschnitten beschrieben.

*Benennen Sie die Patientencharakteristika und patientenrelevanten Endpunkte, die in den relevanten Studien erhoben wurden. Begründen Sie, wenn Sie von den oben benannten Vorgaben abgewichen sind. Beschreiben Sie für jeden Endpunkt, warum Sie ihn als patientenrelevant einstufen, und machen Sie Angaben zur Validität des Endpunkts (z. B. zur Validierung der eingesetzten Fragebögen). Geben Sie für den jeweiligen Endpunkt an, ob unterschiedliche Operationalisierungen innerhalb der Studien und zwischen den Studien verwendet wurden. Benennen Sie die für die Bewertung herangezogene(n) Operationalisierung(en) und begründen Sie die Auswahl. Beachten Sie bei der Berücksichtigung von Surrogatendpunkten Abschnitt 4.5.4.*

*Sofern zur Berechnung von Ergebnissen von Standardverfahren und –software abgewichen wird (insbesondere beim Einsatz spezieller Software oder individueller Programmierung), sind die Berechnungsschritte und ggf. verwendete Software explizit abzubilden. Insbesondere der Programmcode ist in lesbarer Form anzugeben.*

---

<sup>4</sup> Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.

**UX023-CL303****Patientencharakteristika (siehe Abschnitt 4.3.1.2.1):****Allgemeine Patientencharakteristika**

- Alter
- Geschlecht
- Ethnie
- Körpergröße, Gewicht, BMI

**Krankheitsspezifische Charakteristika**

- Auftreten erster XLH-Symptome
- Schmerz
- Zeitpunkt der XLH-Diagnose, *PHEX*-Mutation
- XLH-bedingte Folgeerkrankungen und medizinische Eingriffe
- Vorherige Therapien

**Patientenrelevante Endpunkte**

In der Studie UX023-CL303, die für die Nutzenbewertung von Burosumab in der Indikation Erwachsene mit XLH herangezogen wird, wurden verschiedene Aspekte der Morbidität und Sicherheit erhoben. Da es sich bei XLH um eine schwerwiegende chronische Erkrankung handelt, die potenziell nicht tödlich verläuft, wird in der Studie UX023-CL303 die Mortalität anhand der schweren unerwünschten Ereignisse erhoben und deshalb nicht gesondert dargestellt. Eine Übersicht der patientenrelevanten Endpunkte befindet sich in Tabelle 4-7.

Tabelle 4-7: Übersicht über die patientenrelevanten Endpunkte in der Studie UX023-CL303

Endpunktkategorie und Ausprägung	Beispiele für Erhebung/Messinstrument
<b>Mortalität</b>	
Im Rahmen der Sicherheit erhoben	-
<b>Morbidität</b>	
Serumphosphat	<i>Serumphosphat</i> In mg/dl Prozentsatz erreichter Normwert ( $\geq 2,5$ mg/dl)
Gehfähigkeit	<i>Six-Minute-Walk-Test (6-Minuten-Gehdistanz Test, 6MWT)</i> Absolute Gehdistanz (Gesamtgehdistanz in Meter) Relative Gehdistanz (zurückgelegte Strecke ausgedrückt als Prozentsatz der vorhergesagten Werte auf Grundlage normativer Daten)

<b>Endpunktkategorie und Ausprägung</b>	<b>Beispiele für Erhebung/Messinstrument</b>
Steifheit und physische Funktion	<i>Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)</i> <i>WOMAC Physical Functioning (WOMAC Physische Funktion)</i> <i>WOMAC Stiffness (WOMAC Steifheit)</i>
Schmerzerhebung	<i>Brief Pain Inventory (BPI)</i> <i>BPI Worst Pain (BPI Schlimmster Schmerz)</i> <i>BPI Pain Severity (BPI Schmerzintensität)</i> <i>BPI Pain Interference (BPI Schmerzinterferenz)</i>
Fatigue	<i>Brief Fatigue Inventory (BFI)</i> <i>BFI Worst Fatigue (BFI Schlimmste Fatigue)</i> <i>BFI Global Fatigue Score (BFI Globale Fatigue)</i>
Erfassung von Therapieeffekten	<i>Patient Global Impression of Improvement (PGI-I)</i>
<b>Lebensqualität</b>	
Spezifischer Lebensqualitätsfragebogen	- <sup>a</sup>
<b>Sicherheit</b>	
Unerwünschte Ereignisse (UE)	Gesamtrate von Patienten mit Ereignissen
Schwerwiegende unerwünschte Ereignisse (SUE)	Gesamtrate von Patienten mit Ereignissen
Unerwünschte Ereignisse (UE) mit CTCAE-Grad 3 oder höher: Grad 3: Unterbrechung von täglichen Aktivitäten des Patienten (Therapie/Behandlung erforderlich) Grad 4: Sterberisiko Grad 5: Tod	Gesamtrate von Patienten mit Ereignissen
Unerwünschten Ereignisse, die zum Studienabbruch führten	Gesamtrate von Patienten mit Ereignissen
Unerwünschte Ereignisse von besonderem Interesse	Gesamtrate von Patienten mit Ereignissen
a: Die Lebensqualität wurde in der Studie UX023-CL303 nicht erhoben	

**Im Folgenden wird die Patientenrelevanz der einzelnen Endpunkte und die Validität der verwendeten Messinstrumente diskutiert:**

### **Mortalität**

Bei der Mortalität handelt es sich um einen patientenrelevanten Endpunkt, der jedoch in diesem Dossier nicht separat dargestellt wird. Dies beruht darauf, dass die XLH zwar eine schwere chronische Erkrankung ist, die zu lebenslangen Folgeerkrankungen führt, aber nicht direkt tödlich verläuft. Zudem führt die Applikation von Burosumab ebenfalls nicht zu einer erhöhten Sterblichkeitsrate. Aufgrund dessen wird die Mortalität anhand der schweren unerwünschten Ereignisse erhoben.

## **Morbidität**

### ***Serumphosphat-Konzentration (mg/dl) und Prozentsatz erreichter Normwert***

#### *Patientenrelevanz*

Die XLH ist eine seltene genetisch bedingte chronische Multisystemerkrankung mit einem komplexen Krankheitsbild, die sich bereits im frühen Kindesalter manifestiert und zu lebenslangen schweren Beeinträchtigungen führt. Das Hauptkennzeichen dieser Erkrankung stellt ein chronischer Phosphatverlust dar, der durch eine vermehrte Expression des FGF23 hervorgerufen wird. FGF23 inhibiert die Transkription des Natrium-Phosphat-Kotransportersystems (NaPi-IIa und -IIc) und hemmt dadurch die Rückresorption von Phosphat über die proximalen Nierentubuli [18]. Zusätzlich inhibiert es die Synthese von 1,25(OH)<sub>2</sub>-Vitamin D (Calcitriol) und stimuliert dessen Katabolismus, wodurch die Resorption von Kalzium und Phosphat in das Knochenskelett stark reduziert wird und eine verminderte Knochenmineralisierung zur Folge hat. Die Auswirkungen der chronischen Hypophosphatämie sind umfangreich. Neben einer lebenslangen Beeinträchtigung des stetig stattfindenden Prozesses der Knochenremodellierung, führt der chronische Phosphatverlust zu Osteomalazie, Skelettdeformationen, Knochenschmerzen, Frakturen/Pseudofrakturen und Enthesopathien. Des Weiteren leiden die Patienten bereits in jungen Jahren an stark ausgeprägter Arthrose, Fatigue, Steifheit und Muskelschmerzen [19, 20].

Bis zur Zulassung von Burosumab war keine den Pathomechanismus adressierende Therapieoption für Erwachsene vorhanden, die zu einer langfristigen Normalisierung der Serumphosphat-Konzentration, dem primären Therapieziel bei XLH, führt [21]. Trotz langjähriger Einnahme von Phosphat und aktivem Vitamin D, das laut der einzig verfügbaren S1-Leitlinie für Kinder als Therapie empfohlen wird, konnten daher die Folgen der schweren Multisystemerkrankung nicht verhindert werden [3].

Eine Vermeidung des renalen Phosphatverlusts und der damit assoziierten Erhöhung der Serumphosphat-Konzentration führen dagegen zu einer verbesserten Knochenmineralisierung, reduzieren die Ausprägungen der Osteomalazie, vermindern das Auftreten von Frakturen/Pseudofrakturen und reduzieren die damit verbundenen starken Knochen- und Gelenkschmerzen und ausgeprägte Fatigue [22, 23]. Eine niedrige Serumphosphat-Konzentration kann daher durch die Patienten direkt wahrgenommen werden und stellt deshalb bei der Erkrankung XLH gemäß § 2 Satz 3 AM-NutzenV einen validen patientenrelevanten Endpunkt da.

#### *Operationalisierung und Validität*

Die Serumphosphat-Konzentration (mg/dl) wurde sowohl zu Baseline als auch zu Woche 1, 2, 6, 10, 14, 18, 22, 24, 26, 28, 34, 36, 46, 48, 60, 70, 72, 84, 94, 96 und alle 12 Wochen in der Extensionsphase II (Woche 97 bis ≤ 149) sowie nach vorzeitigem Abbruch / Behandlungsende nach einer Fastenzeit von mindestens 8 h über Nacht gemessen.

Zusätzlich zu der Serumphosphat-Konzentration in mg/dl wurde der prozentuale Anteil an Studienteilnehmern bestimmt, die, gemittelt über die verschiedenen Dosiszyklen zwischen

Baseline und Woche 24, eine Serumphosphat-Konzentration von  $\geq 2,5$  mg/dl in der Mitte des Dosiszyklus und zum Ende des Dosiszyklus erreichten. Das Erreichen einer Serumphosphat-Konzentration im unteren Normbereich stellt bei Erwachsenen mit XLH das primäre Therapieziel dar und ermöglicht eine ausgewogene Mineralisierung des Knochens. Potenzielle Risiken, die mit einer Hyperphosphatämie (Serumphosphat-Konzentration  $> 4,5$  mg/dl) einhergehen, werden durch das gezielte Anstreben des unteren Normbereichs dagegen vermieden. Langfristige Hyperphosphatämien tragen, ähnlich wie bereits für die Phosphatsubstitution beschrieben (Modul 3 Abschnitt 3.1.2), zu einer Abnahme des Calcitriolspiegels und einer geringeren intestinalen Kalziumresorption bei. Infolgedessen wird die Sekretion des Parathormons stimuliert und die Freisetzung von Kalzium und Phosphat aus dem Knochen gefördert [24]. Schlussendlich kann dies zu einem sekundären oder tertiären Hyperparathyreoidismus sowie zu einer ektopischen Mineralisierung führen, die sich in Form einer Nephrokalzinose manifestiert. Analog dazu wird auch in den Therapieempfehlungen und in der Fachinformation als Zielwert der untere Normbereich festgelegt [2].

Dosisfindungsstudien zu Burosumab zeigten, dass der höchste Peak ( $c_{max}$ ) der Serumphosphat-Konzentration, und damit auch der angestrebte Zielbereich von  $\geq 2,5$  mg/dl, in Abhängigkeit von der Dosis 8 – 15 Tage nach einer Injektion mit Burosumab erlangt wird [25]. Anschließend ist eine leichte Abnahme der Serumphosphat-Konzentration zu verzeichnen, die bis zum Tag 50 ohne eine weitere Applikation von Burosumab wieder auf den Ausgangswert sinkt. Auf diesen Beobachtungen basiert das in der Studie UX023-CL303 angewendete Dosierungsschema, welches eine Applikation von Burosumab (1 mg/kg KG) alle 28 Tage beinhaltet. Aufgrund dieser pharmakodynamischen Befunde wurde in der Studie UX023-CL303 der Anteil an Studienteilnehmern, die zur Mitte des Dosiszyklus eine Serumphosphat-Konzentration im unteren Normbereich erreichen, als primärer Endpunkt gewählt. Im Gegensatz zu der Messung am Ende des Dosiszyklus ermöglicht die Messung in der Mitte des Dosiszyklus die sensitivere Analyse des Effekts von Burosumab auf die Serumphosphat-Konzentration, da der Serumphosphatspiegel zu diesem Zeitpunkt maximal ist. Die Serumphosphat-Konzentration zur Mitte des Dosiszyklus stellt daher die validere Messung dar.

Aufgrund der fluktuierenden Serumphosphat-Konzentrationen, die mit dem circadianen Rhythmus assoziiert sind, sollten die Blutproben morgens möglichst jeweils zur gleichen Uhrzeit entnommen werden. Die Auswertung erfolgte verblindet durch ein zentrales Labor. Aufgrund dessen ist die Validität dieses patientenrelevanten Endpunkts zu jeder Zeit gegeben.

### ***Gefähigkeit: 6-Minuten-Gehdistanz Test***

#### *Patientenrelevanz*

Die starken Knochendeformationen, die ausgeprägte Arthrose und Osteomalazie mit einhergehenden Knochenschmerzen, Pseudofrakturen und Steifheit führen bei Erwachsenen mit XLH zu einer erheblichen Einschränkung in der körperlichen Belastbarkeit. Zusätzlich treten aufgrund der über das gesamte Leben bestehenden chronischen Hypophosphatämie Kraft- und Gleichgewichtsstörungen auf. Die damit einhergehenden körperlichen Behinderungen führen nicht nur zu einer geringeren körperlichen Belastbarkeit, sondern auch zu massiven Einschränkungen im Alltag. In einer online durchgeführten Interviewstudie, die

18 XLH-Patienten einschloss, berichteten alle Studienteilnehmer von einer deutlichen Einschränkung der physischen Funktion insbesondere über Schwierigkeiten beim Gehen [26]. Eine Verbesserung der körperlichen Belastbarkeit unter der Burosumab-Therapie stellt gemäß § 2 Satz 3 AM-NutzenV somit bei XLH einen validen und patientenrelevanten Endpunkt dar.

#### *Operationalisierung und Validität*

Die Messung der motorischen Funktion ist sehr komplex, da sich diese aus verschiedenen Komponenten zusammensetzt. Fachexperten von der *Osteoarthritis Research Society International* (OARSI) und dem *Outcome Measures in Rheumatology and Clinical Trials* haben deshalb ein Konsensusstatement entwickelt, das eine Zusammenstellung von Tests beinhaltet, die die motorische Funktion von Patienten mit Arthrose in klinischen Studien am besten abbildet [27]. Diese ausgewählten Tests ermöglichen es, den Behandlungserfolg bei Arthrose konsistent und zuverlässig zu messen. Zu dieser Zusammenstellung an validierten Tests zählt der 6MWT, der in der Studie UX023-CL303 verwendet wurde, um den Einfluss der Burosumab-Therapie auf die körperliche Funktion bei Erwachsenen mit XLH zu analysieren. Hierbei muss der Patient auf einer vorab abgemessenen Strecke kontinuierlich sechs Minuten gehen. Anschließend wird die zurückgelegte Gesamtstrecke in Metern gemessen. Um eine möglichst hohe Validität zu gewährleisten, wird der Test von vorab geschultem klinischen Personal (bevorzugt lizenzierte Physiotherapeuten) auf Basis der *American Thoracic Society Guidelines* (ATS) durchgeführt [28]. Vor Beginn der Studie erhielt jedes Prüfzentrum ein klinisches Handbuch (*Clinical Evaluator User Manual*), in dem detaillierte Angaben zum Ablauf und zur Auswertung des 6MWT enthalten waren [29]. Diese Instruktionen sollten einen einheitlichen Ablauf des 6MWT über den gesamten Studienverlauf unabhängig vom Studienzentrum gewährleisten.

Die einheitliche Durchführung wird im Folgenden detailliert beschrieben:

#### *Laufstrecke*

Als Laufstrecke sollte vom Studienzentrum ein schwach frequentierter Flur gewählt werden, der nicht mit einem Teppichboden ausgelegt war und Maße von ca. 30 m Länge und 2,5 m Breite aufwies [29]. Die Laufstrecke wurde vorab mit zwei Pylonen markiert, die in einem Abstand von 20 Metern aufgestellt waren. Die vorgeschriebene Breite des Flurs sollte dabei ein problemloses Wenden um die Pylonen herum gewährleisten, ohne beim Laufen um die Pylonen die Gehgeschwindigkeit der Testpersonen zu beeinflussen. Während des Tests mussten die Studienteilnehmer für sechs Minuten zwischen diesen beiden Pylonen auf- und abgehen, wobei eine Runde aus dem Hinlaufen zur zweiten Pylone und dem Zurücklaufen zum Ausgangspunkt (erste Pylone) bestand (eine Runde = 40 m).

#### *Probelauf*

Zu Trainingszwecken wurde der 6MWT bereits bei der ersten Screeningvisite durchgeführt. Lediglich Studienteilnehmer, die auch an diesem Training teilgenommen hatten, durften während der Studie den 6MWT absolvieren.

### *Durchführung des 6MWT*

Vor der eigentlichen Durchführung des Tests mussten alle Studienteilnehmer analog zu den ATS-Kriterien sich zunächst zehn Minuten auf einem Stuhl ausruhen [28, 29]. Während dieser Zeit wurde der Gesundheitszustand evaluiert und Anweisungen zum Ablauf des Tests gemacht. Studienteilnehmer, die einen Ruhepuls von über 120 Schlägen pro Minute, einen systolischen Blutdruck von über 180 mm Hg oder einen diastolischen Blutdruck von über 100 mm Hg aufwiesen, wurden von dem Test aus Sicherheitsgründen ausgeschlossen. Orthesen und Gehhilfen waren während des Tests gestattet, Rollstühle durften dagegen nicht verwendet werden.

Neben den allgemeinen Testanweisungen demonstrierte das Studienpersonal zusätzlich den Studienteilnehmern die Wendung um die Pylonen. Anschließend wurde der Patient aufgefordert innerhalb von sechs Minuten so schnell wie möglich um die beiden Pylonen zu gehen. Eine Motivation des Studienteilnehmers durch Zurufe während dieser sechs Minuten erfolgte dagegen nicht. Analog zu den ATS-Kriterien wurde lediglich nach jeder Minute dem Studienteilnehmer die verbleibende Zeit mitgeteilt. Sobald die sechs Minuten erreicht waren, wurde der Studienteilnehmer aufgefordert stehen zu bleiben, um die final zurückgelegte Strecke zu messen. Anschließend setzte sich der Studienteilnehmer wieder auf einen Stuhl, sodass nachträgliche Messungen zum Gesundheitszustand des Patienten durchgeführt werden konnten.

### *Berechnung der zurückgelegten Gesamtdistanz*

Jede zurückgelegte Runde (40 Meter), die sowohl aus dem Hin- als auch dem Zurücklaufen zwischen den beiden Pylonen im Abstand von 20 Metern bestand, wurde durch das lizenzierte Studienpersonal im 6MWT Worksheet anhand einer Strichliste festgehalten. Falls die zuletzt angefangene Runde innerhalb der sechs Minuten nicht vollständig beendet werden konnte, wurde die bis dahin zurückgelegte Distanz gemessen und auf den nächsten vollständigen Meter gerundet. Anschließend wurde die zurückgelegte Gesamtdistanz in Metern berechnet und in den CRF eingetragen.

Weitere Details zur Durchführung des Tests und Einweisung der Studienteilnehmer sind dem *Clinical Evaluator User Manual* zu entnehmen [29].

Neben der tatsächlichen Gesamtdistanz, die innerhalb von sechs Minuten zurückgelegt wurde, wurde auch die innerhalb von sechs Minuten zurückgelegte Strecke als Prozentsatz der vorhergesagten Werte auf Grundlage normativer Daten ermittelt (relativer Wert, normiert auf Alter, Geschlecht und Größe). Die vorausgesagte Distanz basiert auf der Publikation von *Gibbons et al.* und beruht auf einem gesunden Patientenkollektiv, das aus 88 Erwachsenen im Alter zwischen 20 und 80 Jahren besteht [30]. In dieser Studie wurden Referenzwerte für die in sechs Minuten zurückgelegte Distanz basierend auf den Altersspannen 20 – 40 Jahre, 41 – 60 Jahre und 61 – 80 Jahren entwickelt. Der 6MWT wurde, soweit möglich, sowohl zu Studienvisite 1 als auch zu Baseline sowie Woche 12, 24, 36, 48, 72, 96 und alle 24 Wochen in der Extensionsphase II (Woche 97 bis  $\leq$  149) durchgeführt.

### ***Steifheit und physische Funktion: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)***

#### *Patientenrelevanz*

Zu einer der häufigsten Folgeerkrankungen, die durch die stark ausgeprägten Skelettdeformationen und die chronische Hypophosphatämie verursacht werden, zählt das frühe Einsetzen von Arthrose bei erwachsenen XLH-Patienten. Die dadurch hervorgerufene Steifheit sowie die Abnahme der Bewegungsfähigkeit und der Alltagsaktivität führen zu einer starken Einschränkung der Patienten im Alltag. Dies wird auch anhand einer online Interviewstudie ersichtlich, an der 18 XLH-Patienten teilgenommen haben [26]. Sowohl die Steifheit als auch eine verminderte physische Funktion zählten hierbei zu den Hauptsymptomen der Erkrankung, die die Studienteilnehmer berichteten. Die hohe Inzidenz von Arthrose zeigt sich auch anhand der Charakteristika der Studienpopulation der Studie UX023-CL303. Bereits 35 % der 18 – 35-Jährigen sind hiervon betroffen, wobei sich die Anzahl bei den 30 – 49-Jährigen mit 74 % mehr als verdoppelt [5]. Aufgrund dessen stellt eine Abnahme der Steifheit und Einschränkung der physischen Funktion, die direkt durch die Erwachsenen mit XLH wahrgenommen werden kann, gemäß § 2 Satz 3 AM-NutzenV einen patientenrelevanten und validen Endpunkt dar.

#### *Operationalisierung und Validität*

Der WOMAC-Fragebogen (WOMAC LK3.1) zählt zu den bei Hüft- und Kniearthrose am häufigsten verwendeten krankheitsspezifischen Messinstrumenten, und wurde durch das Komitee von *Outcome Measures in Rheumatology Arthritis Clinical Trials* (OMERACT) bei Arthrose empfohlen [31, 32]. Des Weiteren wird dieser Fragebogen bei anderen muskuloskelettalen Erkrankungen wie z. B. Rückenschmerzen, rheumatoider Arthritis, systemischer Lupus Erythematodes und Fibromyalgie eingesetzt. Ziel dieses patientenberichteten Fragebogens ist es, spezifisch die durch Arthrose in den unteren Extremitäten hervorgerufenen Dysfunktionen und Schmerzen zu erfassen. Der Fragebogen besteht aus insgesamt 24 Items und unterteilt sich in die 3 Domänen *Steifheit* (2 Fragen), *physische Funktion* (17 Fragen) und *Schmerzen* (5 Fragen). Während die Domäne *WOMAC Steifheit* das Ausmaß der Steifheit nach dem ersten Gehen am Tag sowie im Verlauf des Tages misst, adressiert die Domäne *WOMAC Physische Funktion* u.a. die Treppennutzung, das Aufstehen, Laufen, Sitzen und Haushaltstätigkeiten. Die Domäne *WOMAC Schmerz* evaluiert dagegen den Schmerz während des Gehens, der Treppenbenutzung, im Bett, sitzend oder liegend. Zum Messen des Ausmaßes von Arthrose auf die Steifheit, die physische Funktion und den Schmerz wurde die Likert-Skala von 0 bis 4 verwendet (keine, milde, moderate, schwere und extreme Auswirkung). Dabei werden für jede Domäne die Werte der Antwortmöglichkeiten addiert, sodass sich verschiedene Spannweiten für die einzelnen Domänen ergeben (*Steifheit* 0 – 8 Punkte, *Physische Funktion* 0 – 68 Punkte, *Schmerz* 0 – 20 Punkte). Ein höherer Score indiziert dabei eine höhere Ausprägung an Steifheit, funktioneller Beeinträchtigung sowie Schmerz. Die Scores werden anschließend auf einer metrischen Skala von 0 – 100 (*normalised units*, normierte Einheiten, nu) normiert, wobei 0 den besten Gesundheitszustand reflektiert und 100 den schlechtesten [33]. Der Fragebogen wurde sowohl zu Baseline, Woche 12, 24, 36, 48, 72 und 96 erhoben, sowie alle 24 Wochen in

der Extensionsphase-II (Woche 97 bis  $\leq 149$ ). Bis einschließlich Woche 24 waren die Studienteilnehmer, das Studienpersonal und die Prüfarzte gegenüber der Behandlung verblindet. Der Fragebogen enthält eine detaillierte Anleitung, sodass kein weiteres Training der Studienteilnehmer notwendig war. Der Bezugszeitraum beträgt 48 Stunden.

Die Domäne *WOMAC Schmerz* wurde in der Studie UX023-CL303 zwar erhoben, aber nicht für die Auswertung berücksichtigt, da der Schmerz bereits durch den Endpunkt *Brief Pain Inventory* (BPI) beurteilt wurde. Im Gegensatz zu dem BPI-Fragebogen, der das Ausmaß an Schmerzen im Allgemeinen sowie die Auswirkung von Schmerzen auf Alltagsaktivitäten, die Lebensfreude und auf Beziehungen zu anderen Menschen misst, wird anhand der Subskala *WOMAC Schmerz* den vom Patienten bei fünf spezifischen Aktivitäten wahrgenommenen Schmerz beim Gehen, Sitzen oder Liegen, Stehen, Treppen auf- und -absteigen sowie nachts beim Schlafen gemessen. Das allgemeine Schmerzlevel unabhängig von der Tätigkeit wird dagegen nicht erhoben. Erwachsene XLH-Patienten weisen aufgrund der seit Jahrzehnten bestehenden chronischen Hypophosphatämie zahlreiche Folgeerkrankungen auf, die mit erheblichen Gelenk- und Knochenschmerzen assoziiert sind [19, 20]. Folglich sind diese Patienten tagtäglich Schmerzen ausgesetzt, die aber nicht auf bestimmte Alltagsaktivitäten beschränkt sind. Insbesondere bei persistierenden chronischen Schmerzen komplexer Art, wie für die XLH beschrieben, ist es daher wichtig zwischen dem allgemein empfundenen Schmerzlevel und dem Schmerz zu differenzieren, der nur unter bestimmten Bedingungen/Belastungen entsteht. Durch die Domäne *WOMAC Schmerz*, die lediglich den Schmerz bei spezifischen Tätigkeiten misst (z. B. beim Gehen), wird daher nur ein Teil der XLH-assoziierten Schmerzen erfasst. Im Gegensatz dazu wird durch Verwendung des *BPI*-Fragebogens ein breiteres Spektrum an XLH-assoziierten Schmerzen erhoben und somit auch der Einfluss von Burosumab auf XLH-assoziierte Schmerzen spezifischer analysiert. Aufgrund dieser Vorteile wurde der BPI-Fragebogen für die Schmerzerhebung in der Studie UX023-CL303 bevorzugt herangezogen und die Domäne Schmerz des *WOMAC*-Fragebogens lediglich *post-hoc* ausgewertet.

Klinische Relevanzschwellen (MID) des *WOMAC* sind für Patienten mit Arthrose etabliert, jedoch nicht spezifisch für XLH entwickelt worden. In der Validierungsstudie für *WOMAC* spezifische MIDs wurde für die Domäne *WOMAC Physische Funktion* eine Abnahme des Scores um mindestens 9,3 normierte Einheiten (*normalised units* = nu) und bei der Domäne *WOMAC Steifheit* eine Abnahme des Scores um mindestens 10 nu als klinisch bedeutsame Veränderung (*minimally important difference*) definiert [34, 35]. Hierbei handelt es sich jeweils um die mittlere Änderung der entsprechenden *WOMAC*-Skala, die von Patienten mit Hüft- und Kniearthrose als Verbesserung in der Beeinträchtigung wahrgenommen werden kann. Für das vorliegende Dossier wird jedoch analog zu dem IQWiG Methodenpapier 6.0 ausschließlich eine MID von  $\geq 15$  nu berücksichtigt, die 15 % der Skalenspannweite entspricht [8]. Diese MID wurde ebenfalls vom G-BA im Nutzenbewertungsverfahren 2020-11-01-D-588 zum Ableiten des Zusatznutzens herangezogen [1].

Sowohl die Kontenvalidität als auch andere Voraussetzungen der *Food and Drug Administration (FDA) Guidance* wurden hinsichtlich der Übertragbarkeit auf XLH evaluiert

und validiert. Die Validierung erfolgte anhand eines mehrstufigen Prozesses und basiert auf den verschiedenen Studien zur XLH bei Erwachsenen. Dazu zählt die Phase-3-Studie UX023-CL303, die Phase-2b-Studie UX023-CL203 sowie die Studie KRN23INT-001/002, die Onlinebefragung UX023-CL001 und die Interviewstudie UX7181B [36]. Der mehrstufige Prozess wird im Folgenden detailliert beschrieben.

#### *Inhaltsvalidität*

Im ersten Schritt wurden anhand der Studie UX7181B, die 18 Erwachsene mit XLH einschloss, die relevantesten Symptome der XLH-Erkrankung und die damit assoziierten Einschränkungen der physischen Funktion sowie der Alltagsaktivitäten identifiziert. Zu den am häufigsten berichteten Einschränkungen der physischen Aktivität zählten das Gehen (n = 18), das Sitzen (n = 15), das Stehen (n = 14) und Schlafstörungen (n = 14). Die am häufigsten berichteten XLH-assoziierten Symptome waren dagegen Schmerzen (n = 18), Steifheit (n = 18), Fatigue (n = 15) und Skelettdeformationen (n = 12). Im nächsten Schritt wurde von jedem Studienteilnehmer der *WOMAC*-Fragebogen ausgefüllt. Generell wurde dieser Fragebogen als leicht verständlich empfunden und deckte die häufigsten und wichtigsten Symptome und Einschränkungen der physischen Funktion, die zuvor von den Patienten berichtet wurden, ab. Analog dazu war die *Item-Response*-Verteilung, analysiert anhand der Studien UX023-CL001, UX023-CL203, UX023-CL303 und KRN23-INT-001/002, normalverteilt.

#### *Konstruktvalidität*

Zur Untersuchung der Konstruktvalidität wurde die Korrelation der einzelnen Items mit den entsprechenden Skalen berechnet. Die Mehrheit dieser Items wies eine hohe Korrelation ( $> 0,6$ ) mit der eigenen Skala auf und bestätigt die Eignung der Item-Skalen-Struktur des Fragebogens für die Anwendung bei der XLH. Zur weiteren Analyse der Konstruktvalidität wurde zudem die Sensitivität hinsichtlich der Krankheitsschwere untersucht. Die verschiedenen Gruppen mit unterschiedlicher Krankheitsschwere wurden anhand mehrerer Kriterien wie z. B. der Einnahme von Schmerzmitteln (ja vs. nein), der Schmerzintensität (schwer, mild/moderat), der Gehfähigkeit (350 m zu Baseline ja vs. nein) und der Vorgeschichte von Knochenfrakturen (ja vs. nein) vorab klassifiziert. Für jeden dieser Parameter, außer dem Parameter Vorgeschichte von Knochenfrakturen, zeigte sich sowohl für die Domäne *WOMAC Steifheit* als auch *WOMAC Physische Funktion* ein signifikanter Unterschied zwischen den Gruppen und bestätigt die Eignung und Sensitivität des Fragebogens zur Differenzierung von XLH-Patienten mit unterschiedlicher Krankheitsschwere.

#### *Konvergenzvalidität*

Des Weiteren wurde die Konvergenzvalidität anhand des patientenberichteten Fragebogens *SF-36* und dem *Patient Global Impression of Improvement* (*PGI-S*) sowie dem *6MWT* evaluiert. Die Domäne *WOMAC Physische Funktion* wies generell eine höhere Konvergenzvalidität (*PGI-S*: 0,562; *6MWT*: 0,511) auf, während die Konvergenzvalidität für *WOMAC Steifheit* in Bezug auf den *6MWT* (-0,316) bzw. *PGI-S* (0,343) niedriger war.

### Reliabilität

Das Cronbachs alpha lag für die Domäne *WOMAC Physische Funktion* zwischen 0,82 – 0,97 (KRN23-INT-001/002, UX023-CL203) bzw. 0,950 – 0,952 (UX023-CL303) und indiziert eine hohe interne Konsistenz und Reliabilität. Für die Domäne *WOMAC Steifheit*, die lediglich aus 2 Items besteht, konnte dagegen das Cronbachs alpha nicht bestimmt werden. Da das Cronbachs alpha direkt von der Anzahl an Items abhängt, wäre bei der Verwendung von lediglich 2 Items eine sehr hohe Korrelation notwendig, um einen ausreichend hohen Wert zu erzielen. Aufgrund dessen wird diese Methode erst ab mindestens 3 Items empfohlen.

Die hohe Reliabilität des WOMAC-Fragebogens spiegelt sich auch anhand der Test-Retest-Reliabilität (*WOMAC Steifheit*: ICC = 0,520 – 0,673; *WOMAC Physische Funktion*: ICC = 0,740 – 0,914) basierend auf dem *PGI-I* für alle analysierten Zeitpunkte wider. Unter Verwendung des *6MWT* war dagegen die Test-Retest-Reliabilität für die Domäne *WOMAC Steifheit* niedriger (0,138 – 0,478), die der Domäne *WOMAC Physische Funktion* dagegen ähnlich hoch wie bereits zuvor unter Verwendung des *PGI-I* (Woche 24 bis Woche 48: 0,629 – 0,836).

### Änderungssensitivität

Zusätzlich wurde die Änderungssensitivität mittels des *6MWT* und des *PGI-I* bestimmt. Anhand des Vergleichs mit dem Fragebogen *PGI-I* (UX023-CL303) zeigte sich, dass durch den *WOMAC*-Fragebogen relevante Änderungen im Gesundheitszustand festgestellt werden können (Tabelle 4-8).

Tabelle 4-8: Übersicht der standardisierten Effektstärke (Cohen`s d): die Unterteilung der verschiedenen Gruppen basiert auf dem *PGI-I* (UX023-CL303)

PGI-I				
Erhebungszeitpunkt		Woche 24	Woche 36	Woche 48
<b>WOMAC Steifheit</b>	verbessert	-0,49	-0,97	-1,03
	unverändert	-0,33	-0,39	-0,04
	verschlechtert	0,06	0,59	0,20
<b>WOMAC Physische Funktion</b>	verbessert	-0,45	-0,60	-0,64
	unverändert	-0,12	-0,22	0,01
	verschlechtert	0,22	0,18	-0,09

Lediglich im Vergleich zum *6MWT* konnte mit dem *WOMAC*-Fragebogen (UX023-CL303) nicht zwischen Patienten mit einem unveränderten oder verbesserten Gesundheitszustand unterschieden werden (Tabelle 4-9).

Tabelle 4-9: Übersicht über die standardisierte Effektstärke (Cohen`s d): die Unterteilung der verschiedenen Gruppen basiert auf dem 6MWT (UX023-CL303)

6MWT				
Erhebungszeitpunkt		Woche 24	Woche 36	Woche 48
WOMAC Steifheit	verbessert	-0,16	-0,36	-0,95
	unverändert	-0,38	-0,97	-0,87
	verschlechtert	0,17	0,23	0,17
WOMAC Physische Funktion	verbessert	-0,17	-0,38	-0,74
	unverändert	-0,29	-0,57	-0,52
	verschlechtert	0,16	0,16	0,59

Höchstwahrscheinlich ist dies auf die bereits zuvor beobachtete geringere Korrelation zwischen den beiden Messinstrumenten zurückzuführen. Eine Erklärung hierfür könnte das unterschiedliche Abdecken der Symptome und der Einschränkungen der physischen Funktion durch die beiden Messinstrumente sein. Während der WOMAC-Fragebogen ein breites Spektrum an Symptomen/Einschränkungen erfasst, werden anhand des 6MWT indirekt die Auswirkungen der XLH auf die Gehfähigkeit evaluiert, ohne jedoch zwischen den verschiedenen Einflussfaktoren zu differenzieren.

Zusammenfassend lässt sich feststellen, dass der WOMAC-Fragebogen ein valides Messinstrument darstellt, um den Effekt von Burosumab auf die Folgeerscheinungen der XLH bei Erwachsenen zu evaluieren. Hier ist insbesondere hervorzuheben, dass die Validität des Fragebogens für alle relevanten Dimensionen nachgewiesen werden konnte. Die Validität des WOMAC-Fragebogens zur Messung der Steifheit und physischen Funktion bei der XLH wurde zudem bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 durch den G-BA bestätigt [1, 37].

### **Schmerz: Brief Pain Inventory (BPI)**

#### *Patientenrelevanz*

Chronische Knochen- und Gelenkschmerzen stellen aufgrund der ausgeprägten Knochendeformationen, Osteomalazie, Arthrose, Enthesopathien sowie der häufig auftretenden Frakturen und Pseudofrakturen die Hauptbeschwerden von XLH-Patienten dar und führen zu einer lebenslangen starken Beeinträchtigung im Alltag und der Lebensqualität. Deshalb ist neben dem Erreichen einer Serumphosphat-Konzentration im unteren Normbereich die Schmerzreduktion ein weiteres Therapieziel bei XLH. Da eine Veränderung des BPI direkte Auswirkungen auf das Wohlbefinden von XLH-Patienten hat, handelt es sich hierbei gemäß § 2 Satz 3 AM-NutzenV um einen patientenrelevanten Endpunkt.

#### *Operationalisierung und Validität*

Der Endpunkt Schmerz wurde anhand der Kurzform des *Brief Pain Inventory* (BPI) erhoben, ein patientenberichteter Fragebogen, der den Schmerz innerhalb der letzten 24 Stunden abfragt. Dieser Fragebogen wurde ursprünglich zur Evaluation des Schmerzes von Krebspatienten

entwickelt, ist mittlerweile jedoch auch für nicht-tumorbedingte Schmerzen aufgrund von Arthrose sowie bei Rückenschmerzen und bei XLH etabliert und validiert [36, 38]. In dem Fragebogen wird anhand von vier Items die Schmerzintensität mit einer 11-Punkte-Skala (0 = kein Schmerz bis 10 = Schlimmster Schmerz, den man sich vorstellen kann) ermittelt (Frage 3 – 6: *BPI Schlimmster Schmerz*, *BPI Durchschnittlicher Schmerz*, *BPI Geringster Schmerz* und *BPI Derzeitiger Schmerz*). Zusätzlich wird die *Schmerzinterferenz* anhand von 7 Items adressiert (9A – 9G). Die Auswertung erfolgt ebenfalls anhand einer 11-Punkte-Skala und reicht von 0 (keine Interferenz) bis 10 (stärkste Interferenz). Die *Schmerzinterferenz* bezieht sich hierbei auf die Beeinträchtigung bei der Ausführung von Alltagsaktivitäten (Gehfähigkeit, Schlaf, Stimmung, Lebensqualität, generelle Aktivitäten, Arbeit und Beziehungen zu anderen).

In der Studie UX023-CL303 wurde die Frage 3 *BPI Schlimmster Schmerz* als sekundärer Endpunkt einzeln erhoben, da dies den ursprünglichen Stratifizierungsfaktor darstellt. Die anderen Items wurden dagegen zusätzlich deskriptiv ausgewertet. Die Frage 3 wurde sowohl bei der ersten Studienvisite als auch in den letzten 7 Tagen vor Studienbeginn (Baseline-Wert) anhand eines Patiententagebuchs erhoben. Der Baseline-Wert setzte sich aus dem Mittelwert des Scores der 7-tägigen Messung und des zu Baseline ermittelten Scores zusammen. Die Messungen fanden ebenfalls zu Woche 12, 24, 36, 48, 72 und 96 statt. Zusätzlich wurde die *Schmerzintensität* anhand der Fragen 3 – 6 des *BPI*-Fragebogens bestimmt (schlimmste, geringste, durchschnittliche und aktuelle Intensität). Diese Domäne wurde, wie bereits beschrieben, ebenfalls anhand eines Patiententagebuches in einem Zeitraum von 7 Tagen vor der jeweiligen Messung sowie zu Baseline, Woche 12, 24, 36, 48, 72 und 96 evaluiert. Hierfür wurde für jede der 4 Fragen der Mittelwert aus dem Score der 8-tägigen Messung, die sich aus einer 7-tägigen Messphase vor dem jeweiligen Messpunkt und dem eigentlichen Messpunkt zusammensetzte, gebildet. Die *Schmerzinterferenz* wurde anhand der Fragen 9A – 9G innerhalb der letzten 24 Stunden zu Baseline, Woche 12, 24, 36, 48, 72 und 96 ermittelt und der Mittelwert der einzelnen Scores für jeden Messpunkt gebildet. Die unterschiedliche Erhebung der Domäne *Schmerzintensität* (über einen Zeitraum von 7 Tagen vor der Studienvisite) und *Schmerzinterferenz* (am Tag der Studienvisite) beruht darauf, dass die *Schmerzintensität* im Verlauf der Zeit variieren kann. Ein längerer Beobachtungszeitraum führt somit zu valideren Aussagen und Änderungen können spezifisch wahrgenommen werden. Die *Schmerzinterferenz* ist dagegen weniger stark Schwankungen ausgesetzt, sodass hier eine Erhebung über einen längeren Zeitraum nicht zu weiteren Erkenntnissen geführt hätte.

Zwei Fragen des *BPI*-Fragebogens wurden in der Studie UX023-CL303 nicht erhoben, da diese separat im Patiententagebuch evaluiert wurden. Hierzu zählen die Einnahme von Schmerzmitteln (Frage 7) und die Linderung der Schmerzen durch die Schmerztherapie (Frage 8). Eine Einweisung der Studienteilnehmer oder der auszuwertenden Personen war bei dem *BPI*-Fragebogen nicht notwendig, da dieser selbsterklärend ist. Der Endpunkt *BPI* stellt somit einen validen und patientenrelevanten Endpunkt bei XLH dar.

Klinische Relevanzschwellen (MID) des *BPI* wurden bereits in klinischen Studien anhand von großen Kohorten validiert, sind jedoch auf Patienten mit chronischen Schmerzen verursacht

durch Neuropathien, Herpes Zoster Neuralgien, Fibromyalgien und Arthrose beschränkt. Für XLH liegt bis zum heutigen Tag keine spezifische Validierung vor, aufgrund der stark ausgeprägten chronischen Knochen- und Gelenkschmerzen kann jedoch angenommen werden, dass die MIDs auch auf Erwachsene mit XLH übertragen werden können. Der kleinste relevante Unterschied (*minimal important difference*, MID), der von einem Patienten wahrgenommen werden kann, ist eine Abnahme des Scores *BPI Schlimmster Schmerz* um mindestens 15 %, wobei eine Abnahme um 30 % eine für den Patienten bedeutsame Abnahme des chronischen Schmerzes darstellt. Sowohl der BPI-Fragebogen als auch die in diesem Dossier dargestellten MIDs wurden bereits durch den G-BA im Nutzenbewertungsverfahren 2020-11-01-D-588 als valide angesehen und für die Nutzenbewertung herangezogen [1, 37].

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

#### *Patientenrelevanz*

Neben Skelettdeformationen ist die XLH ebenfalls durch Erkrankungen des Stütz- und Bewegungsapparates gekennzeichnet und führt neben einer beeinträchtigten Muskelkraft zu einer ausgeprägten Fatigue [20]. Infolgedessen leiden die XLH-Patienten an erheblichen Einschränkungen der physischen Aktivität sowie der Lebensqualität und können nur in eingeschränktem Maße einem normalen Alltag bzw. Arbeitsalltag nachgehen [39]. Hinsichtlich der Erkrankung XLH stellt Fatigue somit einen Endpunkt dar, der direkt von den Erwachsenen wahrgenommen werden kann und wird daher gemäß § 2 Satz 3 AM-NutzenV als patientenrelevant klassifiziert.

#### *Operationalisierung und Validität*

Der Fragebogen *Brief Fatigue Inventory (BFI)* ist ein patientenberichteter Fragebogen, der zwei Dimensionen, die *Schwere der Fatigue* sowie den *Einfluss der Fatigue auf das alltägliche Leben* innerhalb der letzten 24 Stunden, misst. Die *Schwere der Fatigue* wird anhand von 3 Items abgefragt und dient der Messung des schwersten, gewöhnlichen und aktuellen Schweregrads, gemessen an jeweils einer Skala von 0 bis 10, wobei ein höherer Wert eine größere Einschränkung reflektiert. Der Einfluss der Fatigue auf das alltägliche Leben wird anhand von 6 Items (Aktivität, Stimmung, Gehfähigkeit, Arbeit, Beziehungen zu anderen und Lebensfreude) ebenfalls mit einer Skala von 0 bis 10 erfasst. Die Schwere der Fatigue wurde sowohl zu Baseline als auch zu Woche 12, 24, 36 und 48 anhand des Patiententagebuchs erhoben. Die Studienteilnehmer wurden hierfür jeweils 8 Tage vor der Studienvisite von einem Studienkoordinator telefonisch an das Ausfüllen des Fragebogens erinnert. Des Weiteren wurde der gesamte Fragebogen zu Baseline, Woche 12, 24, 36, 48, 72 als auch Woche 96 abgefragt. Um das Ausmaß der *Schwere der Fatigue* zu bestimmen, wurde von jeder Frage der Mittelwert aus den Messungen 7 Tage vor Studienvisite sowie dem Tag der Studienvisite gebildet. Der *Einfluss der Fatigue auf das alltägliche Leben* (Frage 4A – F) wurde dagegen ausschließlich am Tag der Studienvisite erhoben. Der Gesamtwert *BFI Globale Fatigue* (Frage 1 – 3 sowie Frage 4A – F) berechnet sich aus dem Mittelwert der 9 Items am jeweiligen Tag der Studienvisite. Die unterschiedliche Erhebung der Domäne *Schwere der Fatigue* (über einen Zeitraum von 7 Tagen vor der Studienvisite) und *Einfluss der Fatigue auf das alltägliche Leben* (am Tag der Studienvisite) beruht darauf, dass die Schwere der Fatigue im Verlauf der Zeit

variieren kann. Ein längerer Beobachtungszeitraum führt somit zu valideren Aussagen und Änderungen können spezifisch wahrgenommen werden. Der *Einfluss der Fatigue auf das alltägliche Leben* ist dagegen weniger stark Schwankungen ausgesetzt, sodass hier eine Erhebung über einen längeren Zeitraum nicht zu weiteren Erkenntnissen geführt hätte.

Ursprünglich wurde dieser Fragebogen zur Messung der Fatigue bei Krebs und der Behandlung von Krebs entwickelt, wurde aber auch bei Erkrankungen des Bewegungsapparats wie z. B. Arthrose und rheumatoider Arthritis eingesetzt [40-43]. Aufgrund der hohen Inzidenz von Arthrose (74 % der Studienteilnehmer der Studie UX023-CL303) bei Erwachsenen mit XLH stellt dieser Fragebogen daher ein valides Messinstrument dar, den Einfluss von XLH und den damit einhergehenden Folgeerkrankungen auf das Ausmaß und den Schweregrad der Fatigue bei XLH-Patienten zu untersuchen [5]. Die Validität des BFI-Fragebogens zur Messung der Fatigue bei XLH-Patienten wurde bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 durch den G-BA bestätigt [1, 37].

### ***Erfassung von Therapieeffekten: Patient Global Impression of Severity/Improvement (PGI-S, PGI-I)***

#### *Patientenrelevanz*

Der *Patient Global Impression of Severity/Improvement (PGI-S, PGI-I)* ist ein patientenberichteter Fragebogen, der die subjektive Wahrnehmung eines Patienten bezüglich des Therapieerfolgs misst. Da dieser Endpunkt die individuelle Wahrnehmung eines Patienten widerspiegelt, handelt es sich hierbei, laut G-BA Kriterien, um einen patientenrelevanten Endpunkt.

#### *Operationalisierung und Validität*

Der *PGI-S*- bzw. *PGI-I*-Fragebogen ist ein bei Patienten mit chronischen Schmerzen häufig in klinischen Studien eingesetztes Messinstrument [44]. Bei dem *PGI-I* wird die Verbesserung des Gesundheitszustandes anhand einer kategorischen 7-Punkte-Skala, die von sehr stark verbessert bis sehr stark verschlechtert reicht, ermittelt. Die subjektive Wahrnehmung der Krankheitsschwere (*PGI-S*) wird dagegen mit einer kategorischen 4-Punkte-Skala ermittelt, die von 1 (normal) bis 4 (schwer) reicht.

Zu Baseline wurde nur der *PGI-S* eingesetzt, um die patientenindividuelle Wahrnehmung der aktuellen Krankheitsschwere zu bestimmen. Zu Woche 12, 24, 36 und 48 erfolgte dagegen die Evaluation anhand des *PGI-I*, um die Verbesserung bzw. Verschlechterung des Gesundheitszustandes während der Behandlung zu analysieren.

### **Sicherheit**

Alle *unerwünschten Ereignisse* wurden gemäß *Medical Dictionary for Regulatory Activities* (MedDRA, v18.1) ab dem Tag der Unterzeichnung der Einwilligungserklärung bis einschließlich 12 Tage nach der letzten Gabe der Prüfsubstanz aufgezeichnet. In die Auswertung wurden jedoch nur *UE* eingeschlossen, die nach der erstmaligen Verabreichung der Prüfsubstanz (Burosumab) oder Placebo auftraten (*Treatment emergent adverse events (TEAE)*), um die Inzidenz, Häufigkeit und Schwere der *UE* sowie einen möglichen

Zusammenhang mit der Behandlung zu bewerten. *UE*, die zwischen dem Zeitpunkt der Einwilligung und der Baseline-Visite auftraten, wurden hierbei nicht berücksichtigt, da diese *UE* in keinem Bezug zur Behandlung mit der Prüfsubstanz standen und die objektive Beurteilung des Sicherheitsprofils in beiden Behandlungsarmen beeinträchtigt hätten.

Bei jeder Studienvisite oder Hausvisite wurden die Studienteilnehmer nach den, seit der letzten Visite neu aufgetretenen oder andauernden *unerwünschten Ereignisse (TEAE)* befragt und fanden sowohl in der Placebo-Gruppe als auch in der Burosumab-Gruppe statt. Zeitgleich wurden auch die Begleitmedikationen sowie zusätzliche Therapien erfasst. Hierbei wurde besonders darauf geachtet, dass die Dosis der Schmerzmedikation über die gesamte Studiendauer konstant gehalten wurde.

#### *Schwere der UE*

Die Einteilung der *UE* erfolgte anhand von *NCI Common Terminology Criteria for Adverse Events (NCI CTCAE)*. Für den Fall, dass eine Einstufung anhand dieser Kriterien nicht möglich war, wurden die *UE* anhand folgender Kriterien als mild, moderat, schwer, lebensbedrohlich oder tödlich eingestuft.

- Mild (Grad 1): Hierzu zählen alle Anzeichen von Symptomen, die leicht tolerierbar sind und nur geringe Reizungen verursachen und zu keiner Einschränkung in der allgemeinen Aktivität des Patienten führen. Zudem ist keine ärztliche Behandlung notwendig, da diese nur vorübergehend auftreten.
- Moderat (Grad 2): *UE* die zu einem geringen Maß an Unwohlsein oder Besorgnis führen und den Patienten im Alltag möglicherweise einschränken und die körperliche Funktion beeinträchtigen, aber durch einfache therapeutische Maßnahmen verbessert werden können.
- Schwer (Grad 3): Ereignisse, die zu einer Unterbrechung der Alltagsaktivitäten sowie Beeinträchtigung der Studienteilnehmer führen und eine systemische Therapie oder anderweitige Behandlung erfordern.
- Lebensbedrohlich (Grad 4): *UE*, bei denen der Studienteilnehmer dem Risiko ausgesetzt wird zu sterben oder *UE*, die zu Behinderungen führen.
- Tod (Grad 5): *UE*, die zum Tod von Studienteilnehmern führen.

#### *Kausalität der UE*

Anhand folgender Kategorien wurde vom Prüfarzt der Zusammenhang zwischen *UE* und der Prüfsubstanz ermittelt:

- Nichtzusammenhängend: *UE*, die ohne jeglichen Zweifel, definitiv nicht mit der Studienmedikation oder dem Verfahren im Zusammenhang stehen.

- Wahrscheinlich nicht zusammenhängend: *UE*, deren Zusammenhang mit der Prüfsubstanz oder dem Verfahren eher zweifelhaft ist.
- Möglicherweise zusammenhängend: *UE*, die möglicherweise mit der Prüfsubstanz oder dem Verfahren in Zusammenhang stehen.
- Mutmaßlich zusammenhängend: *UE*, bei denen ein Zusammenhang mit der Verabreichung der Prüfsubstanz oder dem Verfahren möglich ist.
- Definitiv zusammenhängend: *UE*, die eindeutig mit der Prüfsubstanz oder dem Verfahren assoziiert sind.

### *Schwerwiegende unerwünschte Ereignisse (SUE)*

Zu den *SUE* zählen folgende Ereignisse:

- Tod
- Lebensbedrohliches *UE*
- *UE*, die eine Hospitalisierung erfordern oder einen Krankenhausaufenthalt verlängern
- Persistierende oder signifikante Einschränkung oder Behinderung (wesentliche Beeinträchtigung der normalen Lebensfunktion)
- kongenitale Anomalie/Geburtsfehler

Hospitalisierungen, die bereits vor Studienbeginn geplant waren, wurden nicht als *SUE* eingestuft. Dagegen wurden Hospitalisierungen, die, nachdem die Rekrutierungsphase bereits abgeschlossen war, aber aufgrund von Vorerkrankungen notwendig waren, als *SUE* eingestuft.

Wichtige medizinische Ereignisse, die nicht tödlich oder direkt lebensbedrohlich verliefen, oder eine Hospitalisierung zur Folge hatten, werden als *SUE* angesehen, sobald, basierend auf medizinischen Beurteilungen, diese den Patienten gefährdeten, oder eine medizinische Maßnahme oder Operation benötigt wurde, um die oben aufgeführten *schweren UE* zu vermeiden.

### *UE von besonderem Interesse*

Folgende *UE* von besonderem Interesse waren prädefiniert:

- 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 kein SMQ verfügbar war, wurde das *UE* mit Hilfe einer MedDRA-Suche nach „Mineralisierung“ definiert
- Restless-Legs-Syndrom: definiert durch die PT „Syndrom der ruhelosen Beine“, „Restlessness“, „Akathisia“, „Sinnesstörungen“, „Psychomotorische Hyperaktivität“, „Gliederschmerzen“, „Neuromuskuläre Schmerzen“ und „Formation“

Entsprechend der Vorgaben des G-BA werden im vorliegenden Nutzendossier *TEAE* und *schwerwiegende TEAE* sowohl als Gesamtraten als auch aufgeteilt nach Systemorganklassen (SOC) bzw. *Preferred Terms* (PT) dargestellt. Da durch den Endpunkt *Sicherheit* eine Verringerung von Nebenwirkungen beurteilt werden kann, handelt es sich hierbei um einen validen und patientenrelevanten Endpunkt.

## Verwendete statistische Methoden

### *Auswertungspopulationen Full Analysis Set (FAS)*

Für die Auswertung der Wirksamkeitsendpunkte wurde die *FAS-Population* verwendet, die dem primären Analyseset entspricht und alle Patienten beinhaltet, die während der doppelblinden Phase der Studie mindestens einmal die Prüfsubstanz verabreicht bekommen hatten. Die Auswertung erfolgte entsprechend der randomisierten Behandlung.

### *Auswertungspopulation Safety Analysis Set*

Für die Auswertung der Sicherheitsendpunkte wurde die *Safety-Population* herangezogen, die definiert war als alle randomisierten Patienten mit mindestens einer Dosis der zugeleiteten Studienmedikation. Die Auswertung erfolgte entsprechend der tatsächlichen Behandlung.

### *Multiples Testen*

Um beim multiplen Testen die *Family-Wise Error Rate* (FWER) zu einem Signifikanzniveau von  $\alpha = 0,05$  zu kontrollieren, wurde für die drei sekundären Wirksamkeitsendpunkten die Hochberg-Adjustierung angewendet (*BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion*) unter  $\alpha = 0,05$ . Laut Methodenpapier des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitssystem (IQWiG) ist dies jedoch für die Nutzenbewertung nicht relevant und wird daher in dem vorliegenden Dossier nicht berücksichtigt [8].

### *Auswertungsmethoden*

Die Ergebnisse der Studie UX023-CL303 werden durch adäquate Effektschätzer mit zugehörigem zweiseitigem 95% KI und zweiseitigen p-Werten geeigneter statistischer Tests zu einem Signifikanzniveau von 5 % dargestellt.

### *Analysen mit stetigen Messwerten*

Für stetige Variablen wurden Änderungen zu Baseline anhand von Mittelwerten inklusive Standardabweichung (SD) berechnet und Unterschiede zwischen den Behandlungsarmen

mittels GEE (*generalized estimation equations*) oder MMRM (*mixed model repeated measures*) analysiert.

Bei den GEE-Modellen wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (*BPI Durchschnittlicher Schmerz*) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline-Wert des entsprechenden Endpunkts als Kovariate. Die abhängige Variable war die Änderung zu Baseline des jeweiligen Endpunkts. Als Kovarianzstruktur wurde eine *Compound Symmetry* Matrix verwendet.

Bei den Auswertungen mittels MMRM (trifft nur auf den Endpunkt *WOMAC* zu) wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (*BPI Durchschnittlicher Schmerz*) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline-Wert des entsprechenden Endpunkts als Kovariate. Die abhängige Variable war die Änderung zu Baseline des jeweiligen Endpunkts. Als Kovarianzstruktur wurde eine unstrukturierte Kovarianz (*WOMAC Steifheit*) bzw. bei nicht-konvergenz eine heterogene Toeplitz Kovarianzstruktur (*WOMAC Physische Funktion*) verwendet.

Um die klinische Relevanz des gemessenen Effekts zu bestimmen, wurde entsprechend den Empfehlungen des IQWiG Methodenpapiers die *SMD post-hoc* in Form von Hedges'g berechnet [8]. Als Irrelevanzschwelle wurde je nach Richtung der Effektschätzer +0,2 oder -0,2 gewählt. Lag das 95% KI vollständig oberhalb oder unterhalb dieser Irrelevanzschwelle, wurde von einem klinisch relevanten Effekt ausgegangen.

#### *Analyse mit binären Messwerten*

Für dichotome Zielvariablen wurden, zusätzlich zu den relativen Häufigkeiten, das *Odds Ratio* (OR), das adjustierte Relative Risiko (ARR), das unadjustierte RR und die absolute Risikodifferenz (ARD) inklusive 95% KI und p-Werten bestimmt.

- Das OR wurde mittels der logistischen Regression bestimmt. Bei den Endpunkten *WOMAC Steifheit*, *WOMAC Physische Funktion* und den Auswertungen der unerwünschten Ereignisse wurde die Behandlungsgruppe und der tatsächliche Stratifizierungsfaktor (*BPI durchschnittlicher Schmerz*) als unabhängige Variablen aufgenommen. Bei dem Endpunkt *Serumphosphat – Prozentsatz erreichter Normwert* ( $\geq 2,5$  mg/dl) und *BPI Schlimmster Schmerz* wurden zusätzlich die Altersgruppen als unabhängige Variable aufgenommen. Aufgrund der geringen Anzahl an Studienteilnehmern aus Japan (n = 11) und Südkorea (n = 7) und folglich einer daraus resultierenden geringen Anzahl an Respondern, konnte der Faktor Region nicht, wie ursprünglich im SAP vorgesehen, als unabhängige Variable im logistischen Regressionsmodell berücksichtigt werden.
- Das adjustierte RR und die dazugehörigen Konfidenzintervalle wurden *post-hoc* nach der Methode von Zhang und Yu bestimmt [45]. Hierfür wurde das adjustierte OR sowie

die Grenzen des entsprechenden Konfidenzintervalls mit folgender Formel transformiert:

$$ARR = \frac{OR}{(1 - P_0) + (P_0 \times OR)}$$

OR = *Odds Ratio*

P<sub>0</sub> = Risiko der Kontrollgruppe

Die Adjustierung erfolgte analog wie bereits zuvor für das OR beschrieben: Bei den Endpunkten *WOMAC Steifheit*, *WOMAC Physische Funktion* und den Auswertungen der unerwünschten Ereignisse wurde die Behandlungsgruppe und der tatsächliche Stratifizierungsfaktor (*BPI durchschnittlicher Schmerz*) als unabhängige Variablen aufgenommen. Bei dem Endpunkt *Serumphosphat – Prozentsatz erreichter Normwert* ( $\geq 2,5$  mg/dl) und *BPI Schlimmster Schmerz* wurden zusätzlich die Altersgruppen als unabhängige Variable aufgenommen. Aufgrund der geringen Anzahl an Studienteilnehmern aus Japan (n = 11) und Südkorea (n = 7) und folglich einer daraus resultierenden geringen Anzahl an Respondern, konnte der Faktor Region nicht, wie ursprünglich im SAP vorgesehen, als unabhängige Variable im logistischen Regressionsmodell berücksichtigt werden.

- Die RR und ARD wurden *post-hoc* unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation und der p-Wert mittels Fishers exaktem Test berechnet.

Das OR, das RR und das ARR wurden als Burosumab gegenüber Placebo berechnet. Die ARD wurde als Burosumab minus Placebo berechnet.

### ***Umgang mit fehlenden Werten***

Fehlende Werte wurden, falls nicht anders spezifiziert, auch als fehlend behandelt. Bei den Endpunkten, bei denen die Änderung zu Baseline evaluiert wurde, wurden nur die Studienteilnehmer mit in die Analysen aufgenommen, bei denen sowohl ein Baseline- als auch ein Post-Baseline-Wert vorhanden war. Abhängig von der Anzahl an fehlenden Werten wurde eine Sensitivitätsanalyse durchgeführt und die fehlenden Werte anhand folgender Methoden imputiert:

- *Modified Baseline Observation Carried Forward* (mBOCF): Bei Abbruch aufgrund eines *UE* wurde der schlechteste Wert zwischen BOCF und LOCF verwendet; andernfalls wurde LOCF verwendet.
- *Baseline Observation Carried Forward* (BOCF): Der nicht fehlende Baseline-Wert wird zur Auswertung weitergeführt.

- *Last Observation Carried Forward (LOCF)*: Der letzte nicht fehlende Wert nach Baseline wird für die Auswertung übernommen

#### *WOMAC-Fragebogen*

Bei dem patientenberichteten Fragebogen *WOMAC* wurden außerdem fehlende Daten folgendermaßen definiert:

Falls zwei oder mehr Items der Domäne *WOMAC Schmerz*, beide Items der Domäne *WOMAC Steifheit* oder mindestens vier Items der Domäne *WOMAC Physische Funktion* fehlten, wurde die entsprechende Domäne als fehlend klassifiziert. Andernfalls wurde jedes fehlende Item durch den Mittelwert der anderen Items in der Domäne ersetzt.

#### *BPI- und BFI-Fragebogen*

Falls die Bewertung eines Items fehlte, wurde dieses als fehlend definiert. Bei der Domäne *BPI Schmerzintensität*, dem *BPI Schmerzintensitätsscore* sowie den Items von *BFI Schwere der Fatigue*, die anhand eines Patiententagebuchs analysiert wurden, wurde der Baseline-Wert zu Studienvisite verwendet, falls drei oder mehr der Tagebucheinträge fehlten. Andernfalls wurden der Mittelwert der Scores vor Studienvisite sowie der Score zur Visite verwendet.

#### *Non-Responder*

Bei dem primären Endpunkt der Studie UX023-CL303 *Serumphosphat-Prozentsatz erreichter Normwert* ( $\geq 2,5$  mg/dl) wurden Studienteilnehmer mit fehlenden Werten als *Non-Responder* klassifiziert.

Bei den klinischen Endpunkten (wie z. B. *WOMAC* oder *BPI*) wurden Studienteilnehmer als *Non-Responder* klassifiziert, sofern zum entsprechenden Zeitpunkt keine Daten verfügbar waren oder der Studienteilnehmer die Studie vorzeitig abbrach.

### **4.2.5.3 Meta-Analysen**

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse quantitativ zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z.B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Für Meta-Analysen soll die im Folgenden beschriebene Methodik eingesetzt werden.

Für die statistische Auswertung sollen primär die Ergebnisse aus Intention-to-treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden. Die Meta-Analysen sollen in der Regel auf Basis von Modellen mit zufälligen Effekten nach der Knapp-

Hartung-Methode mit der Paule-Mandel-Methode zur Heterogenitätsschätzung<sup>5</sup> erfolgen. Im Fall von sehr wenigen Studien ist die Heterogenität nicht verlässlich schätzbar. Liegen daher weniger als 5 Studien vor, ist auch die Anwendung eines Modells mit festem Effekt oder eine qualitative Zusammenfassung in Betracht zu ziehen. Kontextabhängig können auch alternative Verfahren wie z. B. Bayes'sche Verfahren oder Methoden aus dem Bereich der generalisierten linearen Modelle in Erwägung gezogen werden. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, sollen diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt werden.

Für kontinuierliche Variablen soll die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges' g, als Effektmaß eingesetzt werden. Bei binären Variablen sollen Meta-Analysen primär sowohl anhand des Odds Ratios als auch des Relativen Risikos durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen soll ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet<sup>6</sup> werden.

Die Effektschätzer und Konfidenzintervalle aus den Studien sollen mittels Forest Plots zusammenfassend dargestellt werden. Anschließend soll die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand geeigneter statistische Maße auf Vorliegen von Heterogenität<sup>7,5</sup> erfolgen. Die Heterogenitätsmaße sind unabhängig von dem Ergebnis der Untersuchung auf Heterogenität immer anzugeben. Ist die Heterogenität der Studienergebnisse nicht bedeutsam (z. B. p-Wert für Heterogenitätsstatistik  $\geq 0,05$ ), soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten Ausnahmefällen gepoolt werden. Außerdem soll untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.2.5.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.2.5.5).

*Beschreiben Sie die für Meta-Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.*

Nicht zutreffend.

#### **4.2.5.4 Sensitivitätsanalysen**

Zur Einschätzung der Robustheit der Ergebnisse sollen Sensitivitätsanalysen hinsichtlich methodischer Faktoren durchgeführt werden. Die methodischen Faktoren bilden sich aus den

---

<sup>5</sup> Veroniki AA, Jackson D, Viechtbauer W, Bender R, Knapp G, Kuss O et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. *Cochrane Database Syst Rev* 2015: 25-27.

<sup>6</sup> Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.

<sup>7</sup> Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-560.

im Rahmen der Informationsbeschaffung und -bewertung getroffenen Entscheidungen, zum Beispiel die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des Effektmaßes. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ soll für Sensitivitätsanalysen verwendet werden.

Das Ergebnis der Sensitivitätsanalysen kann die Einschätzung der Aussagekraft der Nachweise beeinflussen.

*Begründen Sie die durchgeführten Sensitivitätsanalysen oder den Verzicht auf Sensitivitätsanalysen. Beschreiben Sie die für Sensitivitätsanalysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.*

### ***Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn***

In der Studie UX023-CL303 wurden sowohl für den primären Endpunkt *Serumphosphat-Prozentsatz erreichter Normwert ( $\geq 2,5$  mg/dl)* als auch für die drei sekundären Endpunkte *BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion* Sensitivitätsanalysen durchgeführt.

Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 *Brief Pain Inventory (BPI) Schlimmster Schmerz* des patientenberichteten Fragebogens *BPI* und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens *BPI* bestimmt. Die verkürzte Form des Fragebogens enthielt die Fragen 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3. Das für die Stratifizierung verwendete *Interactive web response system (IWRS)* war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*) verwendet wurde (Anhang 4-E 8b, Tabelle 4-10).

Tabelle 4-10: Nummerierung der einzelnen Items des *BPI*-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	Brief Pain Inventory
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	_ <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	_ <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.
9 A – G	_ <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:
<p>a: IWRS: <i>Interactive web response system</i></p> <p>b: für die Stratifizierung wurde eine verkürzte Form des <i>BPI</i>-Fragebogens verwendet, der nur die Items der Domäne Schmerzintensität enthielt.</p>		

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist.

Zusätzlich erfolgten Sensitivitätsanalysen, um die Robustheit der Ergebnisse zu überprüfen. Für den primären Endpunkt *Serumphosphat* wurde der Cochran–Mantel–Haenszel-Test (CMH-Test) nach dem tatsächlichen Stratifizierungsfaktor *BPI Durchschnittlicher Schmerz* ( $\leq 6,0$  oder  $> 6,0$ ) sowie *BPI Stärkster Schmerz* ( $\leq 6,0$  oder  $> 6,0$ ) und der Region adjustiert. Bei den sekundären Endpunkten *WOMAC Steifheit* und *WOMAC Physische Funktion* wurde der tatsächliche Stratifizierungsfaktor *BPI Durchschnittliche Schmerz* ( $\leq 6,0$  oder  $> 6,0$ ) als Kovariate in das *Generalized estimating equations Modell* (GEE-Modell) aufgenommen, um eine mögliche Auswirkung der fehlerhaften Stratifizierung auf die Ergebnisse zu überprüfen. Da *BPI Schlimmster Schmerz* und *BPI Durchschnittlicher Schmerz* sich sehr ähnlich sind und eine hohe Korrelation zueinander hatten (Pearson Korrelationskoeffizient 0,82256) wurde im GEE-Modell für diesen Endpunkt die Kovariate *BPI Durchschnittlicher Schmerz* nicht mit einbezogen. Die Ergebnisse waren den primären Analysen ähnlich und bestätigen, dass die falsche Stratifizierung zu Beginn der Studie keine Auswirkung auf die Ergebnisse hat.

Für die Auswertungen der Endpunkte in der Studie UX023-CL303 wurde die primäre Analyse verwendet, die sich aus allen Studienteilnehmern zusammensetzte, die zumindest eine Dosis der Prüfsubstanz während der doppelblinden und Placebo-kontrollierten Phase der Studie erhielten ( $N_{\text{(Burosumab)}} = 68$ ;  $N_{\text{(Placebo)}} = 66$ ). Bei Endpunkten, die die Veränderung zu Baseline

evaluierten, wurden nur die Studienteilnehmer in die Analysen mit einbezogen, von denen zumindest ein Baseline-Wert und ein Post-Baseline-Wert vorlag. Dies entspricht somit einer adäquaten Umsetzung des *Intention-To-Treat* (ITT)-Prinzips.

### ***Sensitivitätsanalyse aufgrund fehlender Post-Baseline-Werte***

In der Studie UX023-CL303 wurden generell fehlende Werte als fehlend behandelt, für den Fall, dass dies laut SAP und Studienprotokoll nicht anders spezifiziert war. Für die drei sekundären Endpunkte *BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion* wurden zusätzlich Sensitivitätsanalysen durchgeführt, um die Robustheit der Analysen bei fehlenden Werten zu analysieren. Hierfür wurde das Verfahren des *last observation carried forward* (LOCF), *baseline observation carried forward* (BOCF) und *modified baseline observation carried forward* (mBOCF) eingesetzt (siehe Abschnitt 4.2.5.2). Diese Analysen waren für die Endpunkte *BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion* im Vergleich zu den primären Analysen sehr ähnlich und belegen somit, dass fehlende Werte keinen Einfluss auf die Analysen in der Studie haben.

### ***Sensitivitätsanalyse 6MWT***

In der Studie UX023-CL303 wurde die Gehfähigkeit von XLH-Patienten anhand des 6MWT analysiert. Nach der Schließung der Datenbank für die Primäranalyse von Woche 24 (23.03.2017) stellte das Studienpersonal stark abweichende Messwerte bei einem Studienteilnehmer fest. Dieser Studienteilnehmer legte bei der ersten Screeningvisite eine Gesamtgehdistanz von 508 Metern in sechs Minuten zurück, zu Baseline betrug die Gehdistanz dagegen 860 Meter, zu Woche 12 lediglich 492 Meter.

Medizinisch wäre eine Steigerung der Gehdistanz zu Baseline (Steigerung um 352 Meter im Vergleich zur ersten Screeningvisite) gefolgt von einem rapiden Abfall der Gehdistanz zu Woche 12 (Abnahme der Gehdistanz um 368 Meter) nur durch eine erhebliche Verbesserung bzw. Verschlechterung des Gesundheitszustandes zu erklären gewesen, die im *Case Report Form* (CRF) hätte festgehalten werden müssen. Dies war jedoch während der doppelblinden Phase der Studie für diesen Studienteilnehmer nicht der Fall. Auffallend war auch, dass kein weiterer Studienteilnehmer der Studie bis zu diesem Zeitpunkt eine Gesamtgehdistanz von > 800 Meter in sechs Minuten erreichte. Unterstellt man eine Geschwindigkeit beim Gehen von ca. 5 km/h bei gesunden Probanden, so würden diese eine Strecke von 860 Meter in ca. 10,5 min zurücklegen. Umgekehrt betrachtet hätte der Studienteilnehmer eine Gehgeschwindigkeit von 8,6 km/h aufweisen müssen, um eine solche Distanz in sechs Minuten zurückzulegen und wäre, trotz schwerwiegender Folgeerscheinungen der XLH, somit um ca. 70 % schneller als ein gesunder Proband. Dass ein XLH-Patient eine Strecke von 860 Metern in sechs Minuten schafft ist daher nicht plausibel.

Die unrealistische Abweichung der Gehfähigkeit zu Baseline bei diesem Patienten wurde bei der Bereinigung der Datenbank nach dem Schließen der Datenbank (23.03.2017) vom *Core-Operations-Team* der Studie festgestellt. Dieses Team leitete daraufhin eine Überprüfung des auffälligen Werts durch das Studienzentrum ein, von dem der Wert eingetragen worden war. Dabei stellte sich heraus, dass der Studienteilnehmer zu Baseline nicht, wie ursprünglich im

CRF dokumentiert, 860 Meter zurückgelegt hatte, sondern dass hier ein Berechnungsfehler der zurückgelegten Distanz durch das Studienzentrum vorlag und der Patient lediglich eine Distanz von 460 Metern in sechs Minuten erreicht hatte. Gemäß dem im SOP präspezifizierten Vorgehen wurde die Datenbank geöffnet und der fehlerhafte Wert des Studienteilnehmers korrigiert.

Zum Zeitpunkt dieser Korrektur war sowohl das *Core-Operations-Team*, das die Korrektur veranlasste, als auch das Personal des Studienzentrums, das die Korrektur durchführte, gegenüber der Gruppenzuteilung des fraglichen Patienten noch verblindet. Anlass für die Korrektur war einzig das Auffallen des stark abweichenden Wertes im Rahmen des *data cleaning*. Eine ergebnisgesteuerte Berichterstattung kann daher ausgeschlossen werden.

Um einen potenziellen Einfluss der Neuauswertung auf die Aussagesicherheit der Ergebnisse auszuschließen, wurde für das vorliegende Dossier zusätzlich eine Sensitivitätsanalyse durchgeführt, bei der der Studienteilnehmer mit der fehlerhaften Kalkulation des Baseline-Wertes nicht berücksichtigt wurde. Diese Analyse ergab ähnliche Werte wie die korrigierte Analyse und belegt, dass die Neuberechnung keinen substantziellen Einfluss auf die Interpretation der Ergebnisse hat (siehe Abschnitt 4.3.1.3.1.2). Die korrigierte Analyse des 6MWT wurde bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 anerkannt und für die Ableitung des Zusatznutzens von Burosumab herangezogen [1, 46].

#### ***Sensitivitätsanalyse WOMAC MMRM-Analyse (mixed model repeated measures)***

Für die Auswertungen von mit Skalen erhobenen patientenberichteten Endpunkten war in der Studie UX023-CL303 laut SAP das GEE-Modell vorgesehen. Ähnlich wie das MMRM-Modell, berücksichtigt dieses statistische Modell die über den Studienverlauf ermittelten Informationen vollständig. Um den Anforderungen der mittlerweile gültigen Dossievorlage gerecht zu werden und um auszuschließen, dass durch die Verwendung des MMRM-Modells ggf. abweichende Ergebnisse erzielt werden, wurde beispielhaft für den patientenberichteten Endpunkt WOMAC *post-hoc* eine MMRM-Analyse durchgeführt. Diese Ergebnisse waren konsistent zu den Primäranalysen und belegen, dass das GEE-Modell ein adäquates Auswertungsmodell darstellt, das die über den Studienverlauf ermittelten Informationen vollständig berücksichtigt.

#### **4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren**

Die Ergebnisse sollen hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, untersucht werden. Dies können beispielsweise direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen (z. B. die Dosis) sein. Im Gegensatz zu den in Abschnitt 4.2.5.4 beschriebenen methodischen Faktoren für Sensitivitätsanalysen besteht hier das Ziel, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Eine potenzielle Effektmodifikation soll anhand von Homogenitäts- bzw. Interaktionstests oder von Interaktionstermen aus Regressionsanalysen (mit Angabe von entsprechenden Standardfehlern) untersucht werden. Subgruppenanalysen auf der Basis individueller Patientendaten haben in der Regel eine größere Ergebnissicherheit als solche auf Basis von Meta-Regressionen oder Meta-Analysen unter

Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren, sie sind deshalb zu bevorzugen. Es sollen, soweit sinnvoll, folgende Faktoren bezüglich einer möglichen Effektmodifikation berücksichtigt werden:

- Geschlecht
- Alter
- Krankheitsschwere bzw. –stadium
- Zentrums- und Ländereffekte

Sollten sich aus den verfügbaren Informationen Anzeichen für weitere mögliche Effektmodifikatoren ergeben, können diese ebenfalls begründet einbezogen werden. Die Ergebnisse von in Studien a priori geplanten und im Studienprotokoll festgelegten Subgruppenanalysen für patientenrelevante Endpunkte sind immer darzustellen (zu ergänzenden Kriterien zur Darstellung siehe Abschnitt 4.3.1.3.2).

Bei Identifizierung möglicher Effektmodifikatoren kann gegebenenfalls eine Präzisierung der aus den für die Gesamtgruppe beobachteten Effekten abgeleiteten Aussagen erfolgen. Ergebnisse von Subgruppenanalysen können die Identifizierung von Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen unterstützen.

*Benennen Sie die durchgeführten Subgruppenanalysen. Begründen Sie die Wahl von Trennpunkten, wenn quantitative Merkmale kategorisiert werden. Verwenden Sie dabei nach Möglichkeit die in dem jeweiligen Gebiet gebräuchlichen Einteilungen und begründen Sie etwaige Abweichungen. Begründen Sie die durchgeführten Subgruppenanalysen bzw. die Untersuchung von Effektmodifikatoren oder den Verzicht auf solche Analysen. Beschreiben Sie die für diese Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.*

### **Übersicht über die durchgeführten Subgruppenanalysen**

Erster Schritt zur Überprüfung ist der statistische Nachweis unterschiedlicher Effekte zwischen verschiedenen Subgruppen mithilfe eines Interaktions- oder Homogenitätstests. Damit das Ergebnis der Subgruppenanalyse sinnvoll interpretierbar ist, muss einerseits eine ausreichende Anzahl an Patienten in den Subgruppen eingeschlossen und andererseits genügend Ereignisse aufgetreten sein. In Anlehnung an das IQWiG-Methodenpapier 6.0 wird der Interaktionstest daher nur dann interpretiert, wenn folgende Kriterien erfüllt sind [8].

Tabelle 4-11: Kriterien für die Berechnung der Interaktionstests in Anlehnung an das Methodenpapier 6.0 des IQWiG

	Burosumab		Placebo	
	Anzahl Patienten	Anzahl Ereignisse	Anzahl Patienten	Anzahl Ereignisse
Kategorie 1	n <sub>1</sub>	e <sub>1</sub>	n <sub>2</sub>	e <sub>2</sub>
Kategorie 2	n <sub>3</sub>	e <sub>3</sub>	n <sub>4</sub>	e <sub>4</sub>

e: Anzahl von Ereignissen; n: Anzahl von Patienten

- Es müssen mindestens 10 Patienten in jeder Kategorie enthalten sein:  $n_1 + n_2 \geq 10$  und  $n_3 + n_4 \geq 10$ .
- Zusätzlich gilt bei binären Subgruppen, dass in mindestens einer Kategorie mindestens 10 Ereignisse enthalten sein müssen:  $e_1 + e_2 \geq 10$  oder  $e_3 + e_4 \geq 10$ .

Die Ergebnisse der Interaktionstests werden für alle relevanten Endpunkte im Abschnitt 4.3.1.3.2 tabellarisch dargestellt. Für Subgruppenanalysen, bei denen der p-Wert des Interaktionstests  $\geq 0,05$  ist, wird von einem homogenen Effekt über alle Subgruppen ausgegangen. In diesem Fall ist ausschließlich das Ergebnis der Gesamtstudienpopulation relevant, auf eine Darstellung der Ergebnisse der Subgruppenanalyse im Dossier wird verzichtet. Diese Analysen befinden sich daher ausschließlich im Anhang 4-G.

Ein zum Niveau  $\alpha = 0,05$  signifikantes Ergebnis im Interaktionstest wird zunächst als statistischer Beleg für einen Unterschied zwischen den Subgruppen gewertet, jedoch ist zu beachten, dass – wie oben beschrieben – mindestens 10 Patienten in jedem Behandlungsarm enthalten sein müssen bzw. mindestens 10 Ereignisse in einer Kategorie aufgetreten sein müssen.

### *A priori geplante Subgruppenanalysen*

Anhand der Analyse von Effektmodifikationen können Patientenkollektive identifiziert werden, die maßgeblich von einer Behandlung profitieren. In der Studie UX023-CL303 erfolgte dies anhand von präspezifizierten Subgruppenanalysen. Diese sind in Tabelle 4-12 aufgeführt.

Tabelle 4-12: Präspezifizierte Subgruppen

Endpunkt-Kategorie UX023-CL303	Morbidität						Sicherheit
Subgruppen präspezifiziert	Serumphosphat	BPI	WOMAC	BFI	6MWT	PGI-I	

Endpunkt-Kategorie UX023-CL303	Morbidität						Sicherheit
<i>BPI Schlimmster Schmerz</i> ( $\leq 6,0$ vs. $> 6,0$ )	x	x	x	x	x	x	x
<i>BPI Durchschnittlicher Schmerz</i> ( $\leq 6,0$ vs. $> 6,0$ )	x	x	x	x	x	x	x
<i>Region</i> (Nordamerika/EU vs. Japan vs. Südkorea)	x	x	x	x	x	x	x
<i>Geschlecht</i> (weiblich vs. männlich)	x	x	x	x	x	x	x
<i>Ethnie</i> (weiß vs. nicht weiß)	x	x	x	x	x	x	x
<i>Alter</i> ( $\leq 41$ Jahre vs. $> 41$ Jahre)	x	x	x	x	x	x	x
<i>Verwendung von Schmerzmitteln zu Baseline</i> (ja vs. nein)	x	x	x	x	x	x	x
<i>Verwendung von Opioiden zu Baseline</i> (ja vs. nein)	x	x	x	x	x	x	x

Ziel der präspezifizierten Subgruppenanalysen in der Studie UX023-CL303 war es, mögliche Einflussfaktoren auf den Therapieeffekt von Burosumab zu identifizieren. Die *Subgruppen BPI Schlimmster Schmerz* ( $\leq 6,0$  vs.  $> 6,0$ ) und *BPI Durchschnittlicher Schmerz* ( $\leq 6,0$  vs.  $> 6,0$ ) wurden gewählt, um zu evaluieren, ob Patienten mit einer stärkeren bzw. weniger starken Krankheitslast gleichermaßen von einer Burosumab-Behandlung profitieren. Die Subgruppe *Region* wurde aus Gründen von operativen und logistischen Gesichtspunkten gewählt. Die Subgruppen *Verwendung von Schmerzmitteln zu Baseline* (ja vs. nein) und *Verwendung von Opioiden zu Baseline* (ja vs. nein) wurden gewählt, um zu überprüfen, inwiefern die Einnahme von Opioiden bzw. Schmerzmitteln den Therapieeffekt von Burosumab beeinflusst.

#### **Post-hoc durchgeführte Subgruppenanalysen**

Zudem wurden für alle patientenrelevanten Endpunkte die in Tabelle 4-13 aufgeführten Subgruppenanalysen *post-hoc* durchgeführt.

Tabelle 4-13: *Post-hoc* definierte Subgruppen

Endpunkt-Kategorie UX023-CL303	Morbidität						Sicherheit
Subgruppen <i>post-hoc</i> definiert	Serumphosphat	BPI	WOMAC	BFI	6MWT	PGI-I	
<i>WOMAC Score Steifheit zu Baseline</i> ( $\leq 62,5$ vs. $> 62,5$ )	x	x	x	x	x	x	x
<i>WOMAC Score Physische Funktion zu Baseline</i> ( $\leq 47,8$ vs. $> 47,8$ )	x	x	x	x	x	x	x
<i>WOMAC Schmerzscore zu Baseline</i> ( $\leq 50$ vs. $> 50$ )	x	x	x	x	x	x	x
<i>WOMAC Gesamtscore zu Baseline</i> ( $\leq 49,5$ vs. $> 49,5$ )	x	x	x	x	x	x	x
<i>Frakturen zu Baseline</i> (ja vs. nein)	x	x	x	x	x	x	x
<i>6MWT zu Baseline</i> ( $\leq 365$ m vs. $> 365$ m)	x	x	x	x	x	x	x

Ziel der *post-hoc* definierten Subgruppenanalysen war es zu untersuchen, inwiefern Patienten mit einem weniger starken bzw. mit einem stärkeren Ausmaß an Steifheit, Schmerzen oder Einschränkungen in der Gehfähigkeit von einer Burosumab-Behandlung profitieren. Zusätzlich wurde untersucht, ob sich das Vorhandensein von Frakturen zu Baseline auf einen der Endpunkte in der Studie auswirkt.

Zunächst wurden mittels Interaktionstests mögliche Effektunterschiede zwischen den Kategorien einer Subgruppe bewertet. Diese sind für jeden Endpunkt gesammelt in einer Übersichtstabelle im Abschnitt 4.3.1.3.2 dargestellt. Eine bedeutsame Heterogenität wurde ab einem Interaktions-p-Wert  $< 0,05$  angenommen und als Beleg für eine Effektmodifikation gewertet. Lag ein Beleg für eine Interaktion in einem Endpunkt vor, wurde die Subgruppenanalyse für die entsprechende Population in Abschnitt 4.3.1.3.2 dargestellt und diskutiert.

Bei der Durchführung einer großen Anzahl statistischer Tests ist zu beachten, dass bei einer gewissen Anzahl an durchgeführten Tests aufgrund der vorgegebenen Irrtumswahrscheinlichkeit falsch-positive Ergebnissen auftreten. Ein positiver Interaktionstest ( $p < 0,05$ ) wird hier daher nur als Ausgangspunkt zur Überprüfung einer möglichen Effektmodifikation herangezogen. Liegt ein positiver Interaktionstest vor, werden daher zusätzlich folgende Kriterien geprüft.

- Besteht eine medizinisch, biologisch oder physiologisch plausible Erklärung für die mögliche Effektmodifikation?
- Liegen konsistente Ergebnisse zur möglichen Effektmodifikation über mehrere Endpunkte vor?

Außerdem ist zu beachten, dass aufgrund des von der EMA erteilten *Orphan-Drug-Status* und der damit im Zusammenhang stehenden geringen Anzahl der Studienteilnehmer in der Studie eine „geringe Power“ besteht. Dadurch können möglicherweise tatsächlich vorliegende Effekte nicht mit signifikanten Ergebnissen belegt werden. Außerdem handelt es sich bei einigen Subgruppen um *post-hoc* durchgeführte Analysen, die, nach dem Methodenpapier des IQWiG, nicht methodisch korrekt einer Hypothesenprüfung unterzogen werden konnten [8].

#### 4.2.5.6 Indirekte Vergleiche

Zurzeit sind international Methoden in der Entwicklung, um indirekte Vergleiche zu ermöglichen. Es besteht dabei internationaler Konsens, dass Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen Komparator (häufig als nicht adjustierte indirekte Vergleiche bezeichnet) regelhaft keine valide Analysemethoden darstellen<sup>8</sup>. Eine Ausnahme kann das Vorliegen von dramatischen Effekten sein. An Stelle von nicht adjustierten indirekten Vergleichen sollen je nach Datenlage einfache

<sup>8</sup> Bender R, Schwenke C, Schmoor C, Hauschke D. Stellenwert von Ergebnissen aus indirekten Vergleichen - Gemeinsame Stellungnahme von IQWiG, GMDS und IBS-DR [online]. [Zugriff: 31.10.2016]. URL: [http://www.gmds.de/pdf/publikationen/stellungnahmen/120202\\_IQWIG\\_GMDS\\_IBS\\_DR.pdf](http://www.gmds.de/pdf/publikationen/stellungnahmen/120202_IQWIG_GMDS_IBS_DR.pdf).

adjustierte indirekte Vergleiche<sup>9</sup> oder komplexere Netzwerk-Meta-Analysen (auch als „Mixed Treatment Comparison [MTC] Meta-Analysen“ oder „Multiple Treatment Meta-Analysen“ bezeichnet) für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche berechnet werden. Aktuelle Verfahren wurden beispielsweise von Lu und Ades (2004)<sup>10</sup> und Rücker (2012)<sup>11</sup> vorgestellt.

Alle Verfahren für indirekte Vergleiche gehen im Prinzip von den gleichen zentralen Annahmen aus. Hierbei handelt es sich um die Annahmen der Ähnlichkeit der eingeschlossenen Studien, der Homogenität der paarweisen Vergleiche und der Konsistenz zwischen direkter und indirekter Evidenz innerhalb des zu analysierenden Netzwerkes. Als Inkonsistenz wird dabei die Diskrepanz zwischen dem Ergebnis eines direkten und eines oder mehreren indirekten Vergleichen verstanden, die nicht mehr nur durch Zufallsfehler oder Heterogenität erklärbar ist<sup>12</sup>.

Das Ergebnis eines indirekten Vergleichs kann maßgeblich von der Auswahl des Brückenkomparators bzw. der Brückenkomparatoren abhängen. Als Brückenkomparatoren sind dabei insbesondere Interventionen zu berücksichtigen, für die sowohl zum bewertenden Arzneimittel als auch zur zweckmäßigen Vergleichstherapie mindestens eine direkt vergleichende Studie vorliegt (Brückenkomparatoren ersten Grades). Insgesamt ist es notwendig, die zugrunde liegende Methodik für alle relevanten Endpunkte genau und reproduzierbar zu beschreiben und die zentralen Annahmen zu untersuchen<sup>13, 14, 15</sup>

*Beschreiben Sie detailliert und vollständig die zugrunde liegende Methodik des indirekten Vergleichs. Dabei sind mindestens folgende Angaben notwendig:*

- *Benennung aller potentiellen Brückenkomparatoren ersten Grades und ggf. Begründung für die Auswahl.*

---

<sup>9</sup> Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683-691.

<sup>10</sup> Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.

<sup>11</sup> Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012; 3(4): 312-324.

<sup>12</sup> Schöttker B, Lüthmann D, Boukhemair D, Raspe H. Indirekte Vergleiche von Therapieverfahren. Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.

<sup>13</sup> Song F, Loke YK, Walsh T, Glenn AM, Eastwood AJ, Altman DJ. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; 338: b1147.

<sup>14</sup> Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study *BMJ* 2011; 343 :d4909

<sup>15</sup> Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods* 2013; 4(4): 291-323.

- *Genauere Spezifikation des statistischen Modells inklusive aller Modellannahmen. Bei Verwendung eines Bayes'schen Modells sind dabei auch die angenommenen A-priori-Verteilungen (falls informative Verteilungen verwendet werden, mit Begründung), die Anzahl der Markov-Ketten, die Art der Untersuchung der Konvergenz der Markov-Ketten und deren Startwerte und Länge zu spezifizieren.*
- *Art der Prüfung der Ähnlichkeit der eingeschlossenen Studien.*
- *Art der Prüfung der Homogenität der Ergebnisse direkter paarweiser Vergleiche.*
- *Art der Prüfung der Konsistenzannahme im Netzwerk.*
- *Bilden Sie den Code des Computerprogramms inklusive der einzulesenden Daten in lesbarer Form ab und geben Sie an, welche Software Sie zur Berechnung eingesetzt haben (ggf. inklusive Spezifizierung von Modulen, Prozeduren, Packages etc.; siehe auch Modul 5 zur Ablage des Programmcodes).*
- *Art und Umfang von Sensitivitätsanalysen.*

Nicht zutreffend.

### 4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen

In den nachfolgenden Abschnitten sind die Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen zu beschreiben. Abschnitt 4.3.1 enthält dabei die Ergebnisse aus randomisierten kontrollierten Studien, die mit dem zu bewertenden Arzneimittel durchgeführt wurden (Evidenzstufen Ia/Ib).

Abschnitt 4.3.2 enthält weitere Unterlagen anderer Evidenzstufen, sofern diese aus Sicht des pharmazeutischen Unternehmers zum Nachweis des Zusatznutzens erforderlich sind. Diese Unterlagen teilen sich wie folgt auf:

- Randomisierte, kontrollierte Studien für einen indirekten Vergleich mit der zweckmäßigen Vergleichstherapie, sofern keine direkten Vergleichsstudien mit der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen (Abschnitt 4.3.2.1)
- Nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2)
- Weitere Untersuchungen (Abschnitt 4.3.2.3)

Falls für die Bewertung des Zusatznutzens mehrere Komparatoren (z.B. Wirkstoffe) herangezogen werden, sind die Aussagen zum Zusatznutzen primär gegenüber der Gesamtheit der gewählten Komparatoren durchzuführen (z. B. basierend auf Meta-Analysen unter gemeinsamer Betrachtung aller direkt vergleichender Studien). Spezifische methodische Argumente, die gegen eine gemeinsame Analyse sprechen (z. B. statistische oder inhaltliche Heterogenität), sind davon unbenommen. Eine zusammenfassende Aussage zum Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie ist in jedem Fall erforderlich.

#### 4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel

##### 4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel

###### 4.3.1.1.1 Studien des pharmazeutischen Unternehmers

*Nachfolgend sollen alle Studien (RCT), die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie alle Studien (RCT), für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, benannt werden. Beachten Sie dabei folgende Konkretisierungen:*

- *Es sollen alle RCT, die der Zulassungsbehörde im Zulassungsdossier übermittelt wurden und deren Studienberichte im Abschnitt 5.3.5 des Zulassungsdossiers enthalten sind, aufgeführt werden. Darüber hinaus sollen alle RCT, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, aufgeführt werden.*

- Benennen Sie in der nachfolgenden Tabelle nur solche RCT, die ganz oder teilweise innerhalb des in diesem Dokument beschriebenen Anwendungsgebiets durchgeführt wurden. Fügen Sie dabei für jede Studie eine neue Zeile ein.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Angabe „Zulassungsstudie ja/nein“, Angabe über die Beteiligung (Sponsor ja/nein), Studienstatus (abgeschlossen, abgebrochen, laufend), Studiendauer, Angabe zu geplanten und durchgeführten Datenschnitten und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-14: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer ggf. Datenschnitt	Therapiearme
UX023-CL303	ja	ja	abgeschlossen	<p><b><u>Doppelblinde Placebo-kontrollierte Phase<sup>a</sup>:</u></b> 24 Wochen</p> <p><b><u>Behandlungsfortsetzungsphase (offen, Wechsel von Placebo in die Burosumab-Gruppe):</u></b> 24 Wochen (Woche 25 – 48)</p> <p><b><u>Extensions-Phase-I:</u></b> 48 Wochen (Woche 49 – 96, freiwillig)</p> <p><b><u>Extensions-Phase-II:</u></b> ≤ 53 Wochen (Woche 97 – 149, freiwillig, nur Studienteilnehmer aus den USA)</p>	<p><b><u>Doppelblinde Phase (Woche 0 – 24):</u></b> <u>Intervention:</u> Burosumab <u>Kontrolle:</u> Placebo</p> <p><b><u>Offene Phase (ab Woche 25)</u></b> <u>Intervention:</u> Burosumab</p>
a: Für die Wirksamkeitsanalysen werden in der Nutzenbewertung die Wochen 0 – 24 herangezogen, da es sich hierbei um die randomisierte doppelblinde, Placebo-kontrollierte Phase der Studie UX023-CL303 handelt.					

Geben Sie an, welchen Stand die Information in Tabelle 4-14 hat, d. h. zu welchem Datum der Studienstatus abgebildet wird. Das Datum des Studienstatus soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Der Studienstatus wurde am 02.11.2021 aktualisiert.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-14 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-15: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
Nicht zutreffend.	

#### 4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche

*Beschreiben Sie nachfolgend das Ergebnis der bibliografischen Literaturrecherche. Illustrieren Sie den Selektionsprozess und das Ergebnis der Selektion mit einem Flussdiagramm. Geben Sie dabei an, wie viele Treffer sich insgesamt (d. h. über alle durchsuchten Datenbanken) aus der bibliografischen Literaturrecherche ergeben haben, wie viele Treffer sich nach Entfernung von Dubletten ergeben haben, wie viele Treffer nach Sichtung von Titel und, sofern vorhanden, Abstract als nicht relevant angesehen wurden, wie viele Treffer im Volltext gesichtet wurden, wie viele der im Volltext gesichteten Treffer nicht relevant waren (mit Angabe der Ausschlussgründe) und wie viele relevante Treffer verblieben. Geben Sie zu den relevanten Treffern an, wie vielen Einzelstudien diese zuzuordnen sind. Listen Sie die im Volltext gesichteten und ausgeschlossenen Dokumente unter Nennung des Ausschlussgrunds in Anhang 4-C.*

*[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]*

*Geben Sie im Flussdiagramm auch das Datum der Recherche an. Die Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.*

*Orientieren Sie sich bei der Erstellung des Flussdiagramms an dem nachfolgenden Beispiel.*

Im Rahmen der bibliographischen Literaturrecherche wurden folgende Datenbanken durchsucht: *MEDLINE*, *EMBASE* und das *Cochrane Central Register of Clinical Trials*. Die Suche wurde am 02.11.2021 durchgeführt. Die Suchkriterien sind im Abschnitt 4.2.2 dargestellt.

Insgesamt wurden 155 Treffer identifiziert. Nach Ausschluss von Duplikaten blieben noch 75 Treffer übrig. Die Auswahl relevanter Studien aus diesem Pool erfolgte durch zwei unabhängige Reviewer anhand der in Abschnitt 4.2.2 beschriebenen Kriterien. Treffer in Studienregistern wurden dabei nicht berücksichtigt, da diese bereits über die Studienregistersuche abgedeckt sind (Abschnitt 4.3.1.1.3). Durch dieses systematische Vorgehen wurden drei relevante Treffer für die Nutzenbewertung von Burosumab identifiziert (*A Phase 3 randomized, double-blind, placebo-controlled study investigating the efficacy and safety of Burosumab, an anti-FGF23 antibody, in adult X-Linked Hypophosphatemia (XLH)*) (*J Bone Miner Res* (2018) 33(8):1383-1393. doi: 10.1002/jbmr.3475), *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* (*Calcified tissue international*; 105(3):271-84), *Burosumab treatment in adults with X-linked*

*hypophosphataemia: 96-week patient-reported outcomes and ambulatory function from a randomised phase 3 trial and open-label extension.* (RMD open; 7(3): e001714)) [4, 23, 47]. Hierbei handelt es sich um die Publikationen der Studienergebnisse der Zulassungsstudie UX023-CL303. Der Ablauf der Recherche ist im Flussdiagramm Abbildung 1 dargestellt.

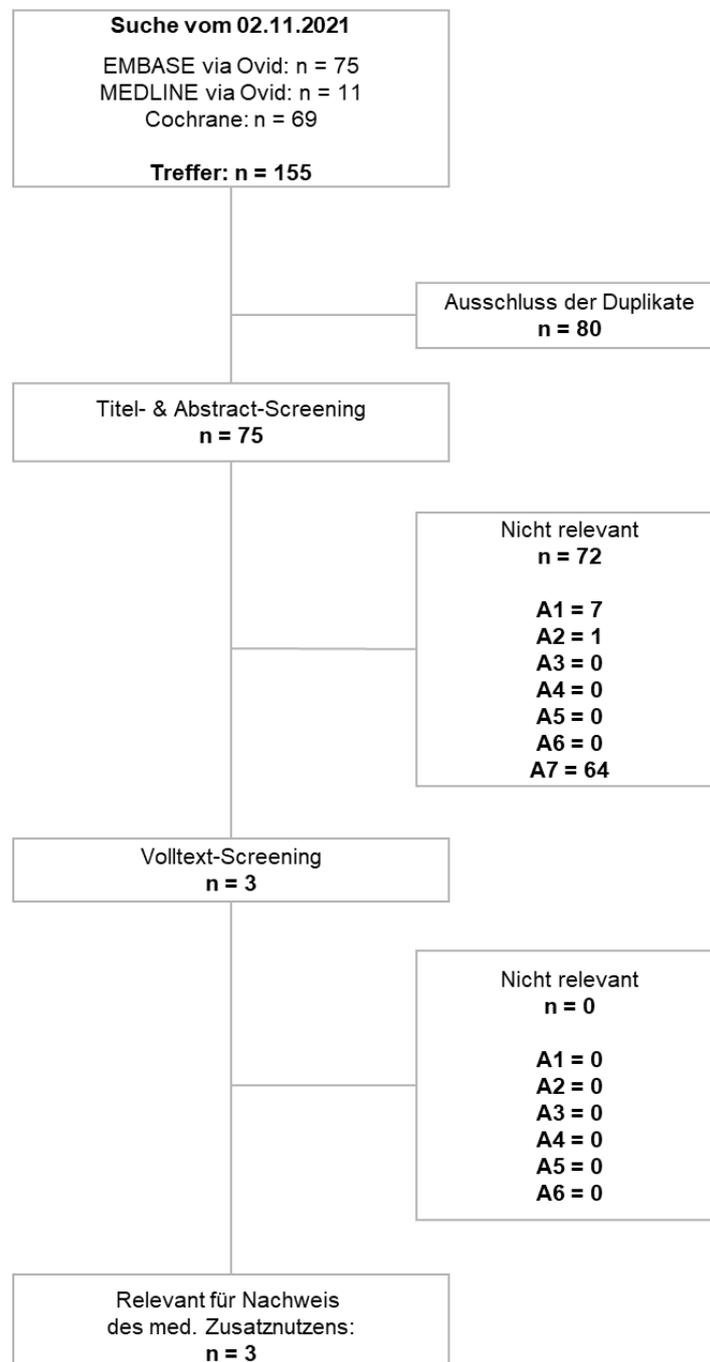


Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel

#### 4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern/ Studienergebnisdatenbanken identifiziert wurden. Geben Sie dabei an, in welchem Studienregister / Studienergebnisdatenbank die Studie identifiziert wurde und welche Dokumente dort zur Studie jeweils hinterlegt sind (z. B. Studienregistereintrag, Bericht über Studienergebnisse etc.). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-14) und ob die Studie auch durch die bibliografische Literaturrecherche identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein. Listen Sie die ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-D.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei Ihren Angaben an der beispielhaften ersten Tabellenzeile.

Tabelle 4-16: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern / Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel

Studie	Identifikationsorte (Name des Studienregisters/ der Studienergebnisdatenbank und Angabe der Zitate <sup>a</sup> )	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)	Status (abgeschlossen/ abgebrochen/ laufend)
<b>UX023-CL303</b>	Clinicaltrials.gov NCT02526160 [10]  WHO ICTRP 2014-005529-11 [12]  EU Clinical Trials Registry 2014-005529-11 [11]  EMA <i>clinical data</i> <sup>b</sup>  Arzneimittel-Informationssystem (AMIce) <sup>b</sup>	ja	ja	abgeschlossen
<p>a: Zitat des Studienregistereintrags, die Studienregisternummer (NCT-Nummer, EudraCT-Nummer) sowie, falls vorhanden, der im Studienregister/in der Studienergebnisdatenbank aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p> <p>b: Die Suche unter EMA <i>clinical data</i> und im Arzneimittel-Informationssystem (AMIce) am 02.11.2021 ergab keinen Treffer.</p>				

Geben Sie an, welchen Stand die Information in Tabelle 4-16 hat, d. h. zu welchem Datum die Recherche durchgeführt wurde. Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Der Stand der Information in Tabelle 4-16 ist der 02.11.2021.

#### 4.3.1.1.4 Studien aus der Suche auf der Internetseite des G-BA

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Sichtung der Internetseite des G-BA identifiziert wurden. Geben Sie dabei an, welche Dokumente dort hinterlegt sind (z. B. Dossier eines anderen pharmazeutischen Unternehmers, IQWiG Nutzenbewertung). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-14) und ob die Studie auch durch die bibliografische Literaturrecherche bzw. Suche in Studienregistern/Studienergebnisdatenbank identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei Ihren Angaben an der beispielhaften ersten Tabellenzeile.

Tabelle 4-17: Relevante Studien aus der Suche auf der Internetseite des G-BA – RCT mit dem zu bewertenden Arzneimittel

Studie	Relevante Quellen <sup>a</sup>	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)	Studie durch Suche in Studienregistern / Studienergebnisdatenbanken identifiziert (ja/nein)
UX023-CL303 <sup>b</sup>	Vorgangsnummer 2020-11-01-D-588 [1, 37, 46, 48, 49]	ja	ja	ja

a: Quellen aus der Suche auf der Internetseite des G-BA  
b: Burosumab wurde vor Überschreitung der Umsatzgrenze von 50 Millionen Euro bereits vom G-BA mit den gleichen Daten Nutzen bewertet.

Geben Sie an, welchen Stand die Information in Tabelle 4-17 hat, d. h. zu welchem Datum die Recherche durchgeführt wurde. Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Suche erfolgte am 09.11.2021. Bei der identifizierten Studie handelt es sich um die Studie UX023-CL303 die bereits nutzenbewertet wurde und im vorliegenden Dossier für die Wiedereinreichung nach Überschreitung der Umsatzgrenze von 50 Millionen Euro erneut für die Bewertung des Zusatznutzens von Burosumab herangezogen wird (Vorgangsnummer 2020-11-01-D-588).

#### 4.3.1.1.5 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel

Benennen Sie in der nachfolgenden Tabelle den aus den verschiedenen Suchschritten (Abschnitte 4.3.1.1.1, 4.3.1.1.2, 4.3.1.1.3 und 4.3.1.1.4) resultierenden Pool relevanter Studien (exklusive laufender Studien) für das zu bewertende Arzneimittel, auch im direkten Vergleich

zur zweckmäßigen Vergleichstherapie. Führen Sie außerdem alle relevanten Studien einschließlich aller verfügbaren Quellen in Abschnitt 4.3.1.4 auf. Alle durch die vorhergehenden Schritte identifizierten und in der Tabelle genannten Quellen der relevanten Studien sollen für die Bewertung dieser Studien herangezogen werden.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Quellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Hierbei sollen die Studien durch Zwischenzeilenüberschriften ggf. sinnvoll angeordnet werden, beispielsweise nach Therapieschema (Akut-/Langzeitstudien) und jeweils separat nach Art der Kontrolle (Placebo, zweckmäßige Vergleichstherapie, beides). Sollten Sie eine Strukturierung des Studienpools vornehmen, berücksichtigen Sie diese auch in den weiteren Tabellen in Modul 4.

Tabelle 4-18: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	Studienkategorie			verfügbare Quellen <sup>a</sup>		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie <sup>b</sup> (ja/nein)	Studie Dritter (ja/nein)	Studienberichte Studienprotokoll SAP (ja/nein [Zitat])	Register-einträge <sup>c</sup> (ja/nein [Zitat])	Publikation und sonstige Quellen <sup>d</sup> (ja/nein [Zitat])
<b>ggf. Zwischenüberschrift zur Strukturierung des Studienpools</b>						
<b>Placebo-kontrolliert<sup>e</sup></b>						
<b>UX023-CL303</b>	ja	ja	nein	ja [5, 50, 51]	ja [10-12]	ja [1, 4, 23, 37, 46-49]
<b>aktivkontrolliert</b>						
Nicht zutreffend.						
<p>a: Bei Angabe „ja“ sind jeweils die Zitate der Quelle(n) (z. B. Publikationen, Studienberichte, Studienregistereinträge) mit anzugeben, und zwar als Verweis auf die in Abschnitt 4.6 genannte Referenzliste. Darüber hinaus ist darauf zu achten, dass alle Quellen, auf die in dieser Tabelle verwiesen wird, auch in Abschnitt 4.3.1.4 (Liste der eingeschlossenen Studien) aufgeführt werden.</p> <p>b: Studie, für die der Unternehmer Sponsor war.</p> <p>c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p> <p>d: Sonstige Quellen: Dokumente aus der Suche auf der Internetseite des G-BA.</p> <p>e: Placebo in Kombination mit einer Begleittherapie wird als geeignete Umsetzung von „best supportive care“ angesehen. Eine ausführliche Begründung für die Wahl der zVT ist in Modul 3 Abschnitt 3.1.2 zu finden.</p>						

### 4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

#### 4.3.1.2.1 Studiendesign und Studienpopulationen

Beschreiben Sie das Studiendesign und die Studienpopulation der in die Bewertung eingeschlossenen Studien mindestens mit den Informationen in den folgenden Tabellen. Falls Teilpopulationen berücksichtigt werden, ist die Charakterisierung der Studienpopulation auch

*für diese Teilpopulation durchzuführen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Geben Sie bei den Datenschnitten auch den Anlass des Datenschnittes an. Fügen Sie für jede Studie eine neue Zeile ein.*

*Weitere Informationen zu Studiendesign, Studienmethodik und Studienverlauf sind in Anhang 4-E zu hinterlegen.*

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-19: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

Studie	Studiendesign <RCT, doppelblind/einfach, verblindet/offen, parallel/cross-over etc.>	Population <relevante Charakteristika, z. B. Schweregrad>	Interventionen (Zahl der randomisierten Patienten)	Studiendauer/ Datenschnitte <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
UX023- CL303	Randomisierte, doppelblinde, Placebo-kontrollierte multizentrische Studie der Phase III	Erwachsene mit XLH (18 – 65 Jahre)	<p><b>Burosumab:</b> (1 mg/kg KG) subkutan alle 4 Wochen (n = 68)</p> <p><b>Placebo:</b> subkutan alle 4 Wochen (n = 66) (entsprach der gleichen Zusammensetzung wie die Burosumab- Lösung, jedoch ohne die aktive Substanz)</p> <p>In der Behandlungs- fortsetzungsphase ab Woche 25 erhielten alle Studienteil- nehmer, unabhängig von der Randomisierung, Burosumab wie oben beschrieben.</p>	<p><b>Doppelblinde Placebo- kontrollierte Phase:</b> 24 Wochen (Placebo: 66 Studienteilnehmer, Burosumab: 68 Studienteilnehmer)</p> <p><b>Behandlungsfort- setzungsphase:</b> 24 Wochen (Placebo: 66 Studienteilnehmer, Burosumab: 67 Studienteilnehmer)</p> <p><b>Extensionsphase-I:</b> 48 Wochen (Placebo: 63 Studienteilnehmer, Burosumab: 63 Studienteilnehmer)</p> <p><b>Extensionsphase-II:</b> ≤ 53 Wochen (nur Studienteilnehmer aus den USA) (Placebo: 52 Studienteilnehmer, Burosumab: 49 Studienteilnehmer)</p>	<p><b>Studienzentren:</b> USA, Vereinigtes Königreich Großbritannien, Japan, Frankreich, Südkorea, Irland, Italien</p> <p><b>Studienbeginn:</b> 22.10.2015</p> <p><b>Ende Primäranalyse (Woche 24):</b> 22.12.2016</p> <p><b>Ende Behandlungsfort- setzungsphase (Woche 25 – 48):</b> 08.06.2017</p> <p><b>Ende Extensionsphase-I (Woche 49 – 96):</b> 06.12.2018</p> <p><b>Ende Extensionsphase-II (Woche 97 – Woche ≤ 149):</b></p>	<p><b>Primärer Endpunkt:</b> Anteil an Studienteilnehmern, die in der Mitte des Dosisintervalls eine <i>Serumphosphat- Konzentration</i> oberhalb der unteren Grenze des Normbereichs von 2,5 mg/dl erreichten, gemittelt über die Dosiszyklen zwischen Baseline und Woche 24</p> <p><b>Sekundäre Wirksamkeitsendpunk- te:</b> <i>Brief Pain Inventory (BPI) Schlimmster Schmerz Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Steifheit WOMAC Physische Funktion</i></p> <p><b>Weitere sekundäre Wirksamkeits- bzw. explorative Endpunkte:</b></p>

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

	individuell	<p><i>Serumphosphat mg/dl</i>  <i>6-Minuten-Gehdistanz</i>  <i>Test (6MWT)</i>  <i>BPI Schmerz-</i>  <i>interferenz</i>  <i>BPI Schmerz-intensität</i>  <i>Brief Fatigue Inventory</i>  <i>(BFI) Schlimmste</i>  <i>Fatigue, BFI Globale</i>  <i>Fatigue</i>  <i>Patient Global</i>  <i>Impression of</i>  <i>Improvement (PGI-I)</i></p> <p><b><u>Sicherheitsendpunkte:</u></b>  <i>UE</i>  <i>Schwerwiegende UE</i>  <i>Unerwünschte</i>  <i>Ereignisse von</i>  <i>besonderem Interesse</i></p>
--	-------------	--

Tabelle 4-20: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	Placebo	Burosumab	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in- Phase etc.</i>
<p><b>UX023-CL303</b></p>	<p><b><u>Doppelblinde Placebo-kontrollierte Phase</u></b> <b><u>Woche 0 – 24:</u></b></p> <p>Die Studienteilnehmer erhielten alle 4 Wochen Placebo subkutan. Dabei handelte es sich um eine Lösung, die eine identische Zusammensetzung wie die Prüfsubstanz aufwies, jedoch nicht die aktive Substanz enthielt. Placebo wurde entweder in den Bauch, die Beine oder die Oberarme injiziert, wobei ein Gesamtvolumen von 1,5 ml nicht überschritten werden durfte.</p> <p><u>Leitlinien oder einheitliche Therapieempfehlungen:</u></p> <p>Bis zum heutigen Zeitpunkt sind keine Leitlinien oder einheitliche Therapieempfehlungen für Erwachsene mit XLH vorhanden.</p> <p><u>Behandlungsfortsetzungsphase</u> <u>Woche 25 – 48:</u></p> <p>Die Studienteilnehmer erhielten Burosumab alle 4 Wochen subkutan (1 mg/kg KG).</p> <p><u>Extensionsphase-I</u> <u>(Woche 49 – 96):</u></p> <p>Die Studienteilnehmer erhielten, wie bereits oben beschrieben, alle 4 Wochen Burosumab subkutan.</p> <p><u>Extensionsphase-II</u> <u>(Woche 97 – 149):</u></p> <p>Die Studienteilnehmer (nur aus USA) erhielten, wie bereits oben beschrieben, alle 4 Wochen Burosumab subkutan.</p>	<p><b><u>Doppelblinde Placebo-kontrollierte Phase</u></b> <b><u>Woche 0 – 24:</u></b></p> <p>Die Initialdosis von Burosumab betrug 1 mg/kg KG und wurde alle 4 Wochen subkutan verabreicht. Burosumab wurde entweder in das Abdomen, die Beine oder die Oberarme appliziert, wobei eine Höchstdosis von 90 mg nicht überschritten werden durfte und je Injektionsstelle ein Volumen von maximal 1,5 ml verabreicht werden durfte. Eine mögliche Dosis-Anpassung richtete sich nach dem Nüchtern-Serumphosphatwert.</p> <p><u>Behandlungsfortsetzungsphase</u> <u>Woche 25 – 48 (offene Phase):</u></p> <p>Die Studienteilnehmer erhielten, wie bereits oben beschrieben, alle 4 Wochen Burosumab subkutan.</p> <p><u>Extensionsphase-I</u> <u>(Woche 49 – 96):</u></p> <p>Die Studienteilnehmer erhielten, wie bereits oben beschrieben, alle 4 Wochen Burosumab subkutan.</p> <p><u>Extensionsphase-II</u> <u>(Woche 97 – 149):</u></p> <p>Die Studienteilnehmer erhielten, wie bereits oben beschrieben, alle 4 Wochen Burosumab subkutan.</p>	<p>Studienteilnehmer, die vor Studienbeginn eine Substitutionstherapie aus oralem Phosphat und aktivem Vitamin D einnahmen, mussten diese nach dem Unterschreiben der Einwilligungserklärung für die Laufzeit der Studie absetzen. Bei diesen Studienteilnehmern fand die 2. Studienvisite frühestens 14 Tage nach der letzten Einnahme von oralem Phosphat und Vitamin D statt.</p>

Tabelle 4-21: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel

Studie	UX023-CL303	
	Placebo	Burosumab
Gruppe	66	68
N (FAS)	66	68
<b>Allgemeine Charakteristika</b>		
<i>Alter, MW (SD)</i>	38,65 (12,76)	41,29 (11,58)
<i>Weibliches Geschlecht, n (%)</i>	43 (65,20)	44 (64,70)
<i>Männliches Geschlecht, n (%)</i>	23 (34,80)	24 (35,30)
<i>Ethnie I (Asiaten), n (%)</i>	9 (13,60)	12 (17,60)
<i>Ethnie II (Weiße), n (%)</i>	53 (80,30)	55 (80,90)
<i>Ethnie III (Schwarz oder afrikanische Amerikaner), n (%)</i>	3 (4,50)	0 (0)
<i>Ethnie IV (Anders), n (%)</i>	1 (1,50)	1 (1,50)
<i>Ethnizität hispanisch/lateinamerikanisch, n (%)</i>	5 (7,60)	7 (10,30)
<i>Ethnizität nicht-hispanisch/lateinamerikanisch, n (%)</i>	61 (92,40)	61 (89,70)
<i>Gewicht [kg], MW (SD)</i>	71,27 (18,89)	70,06 (19,00)
<i>BMI [kg/m<sup>2</sup>], MW (SD)</i>	30,60 (7,79)	29,98 (7,49)
<i>Größe [cm], MW (SD)</i>	152,69 (11,84)	152,15 (9,49)
<b>Krankheitsbezogene Baseline-Charakteristika</b>		
<i>Jahre seit Auftreten erster XLH-Symptome, MW (SD)</i>	38,44 (13,17)	35,83 (12,39)
<i>Jahre seit Diagnose, MW (SD)</i>	31,36 (15,79)	31,47 (15,59)
<i>PHEX-Mutation<sup>a</sup>, n (%)</i>	65 (98,48)	61 (89,71)
<i>Serumphosphat-Konzentration (mg/dl), MW (SD)</i>	1,92 (0,32)	2,03 (0,30)
<b>Schmerz zu Baseline</b>		
<i>BPI Durchschnittlicher Schmerz ≤ 6,0, MW (SD)</i>	49 (74,20)	46 (67,60)
<i>BPI Durchschnittlicher Schmerz &gt; 6,0, MW (SD)</i>	17 (25,80)	22 (32,40)
<i>BPI Schlimmster Schmerz ≤ 6,0, MW (SD)</i>	23 (34,80)	15 (22,10)
<i>BPI Schlimmster Schmerz &gt; 6,0, MW (SD)</i>	43 (65,20)	53 (77,90)
<b>Biochemische Parameter</b>		
<i>Serum 1,25(OH)<sub>2</sub>D (pg/ml), MW (SD)</i>	33,5 (15,61)	32,4 (12,96)
<i>TmP/GFR (mg/dl), MW (SD)</i>	1,598 (0,37)	1,678 (0,40)
<b>XLH-assoziierte Folgeerkrankungen</b>		
<b><u>Knochenerkrankung</u></b>	63 (95,5)	64 (94,1)
<i>Skelettdeformationen</i>	62 (93,9)	64 (94,1)
<i>Obere Extremitäten</i>	59 (89,4)	58 (85,3)
<i>Untere Extremitäten</i>	55 (83,3)	60 (88,2)
<i>Osteophyten, n (%)</i>	25 (37,9)	29 (42,6)
<i>Enthesopathie, n (%)</i>	65 (98,5)	68 (100)
<i>Arthrose, n (%)</i>	38 (57,6)	47 (69,1)

Studie	UX023-CL303	
	Placebo	Burosumab
<b>Gruppe</b>	<b>66</b>	<b>68</b>
<b>N (FAS)</b>	<b>66</b>	<b>68</b>
<i>Deformität der Füße</i>	2 (3,0)	1 (1,5)
<i>Kraniosynostose</i>	5 (7,6)	5 (7,4)
<i>Genu valgum</i>	10 (15,2)	13 (19,1)
<i>Genu varum</i>	54 (81,8)	46 (67,6)
<i>Innenfehlstellung der Zehen</i>	16 (24,2)	16 (23,5)
<i>Rachitischer Rosenkranz</i>	6 (9,1)	6 (8,8)
<i>Spinalkanalstenose</i>	10 (15,2)	15 (22,1)
<i>Tibiatorsion</i>	15 (22,7)	20 (29,4)
<i>Verdickung der Handgelenke</i>	12 (18,2)	13 (19,1)
<u>Ektopische Mineralisierung</u>		
<i>Nephrolithiasis, n (%)</i>	8 (12,1)	10 (14,7)
<i>Nephrokalzinose, n (%)</i>	5 (7,6)	11 (16,2)
<i>Hyperparathyreoidismus, n (%)</i>	3 (4,5)	4 (5,9)
<u>Zahngesundheit</u>		
<i>Zahnabszesse, n (%)</i>	41 (62,1)	44 (64,7)
<i>Entfernung von bleibenden Zähnen, n (%)</i>	35 (53,0)	43 (63,2)
<i>Wurzelkanal Operation, n (%)</i>	32 (48,5)	42 (61,8)
<i>Exzessiver Karies, n (%)</i>	32 (48,5)	42 (61,8)
<u>Orthopädische Eingriffe</u>		
<i>Osteotomie, n (%)</i>	38 (57,6)	38 (55,9)
<i>Kraniotomie/Kraniektomie, n (%)</i>	1 (1,5)	0 (0)
<i>Externe Fixierung, n (%)</i>	7 (10,6)	8 (11,8)
<i>Fixierung der Fraktur mit intramedullärem Nagel/Stab, n (%)</i>	6 (9,1)	10 (14,7)
<i>Fixierung der Fraktur mit Platten und Schrauben, n (%)</i>	8 (12,1)	11 (16,2)
<i>Hüftgelenkersatz, n (%)</i>	3 (4,5)	4 (5,9)
<i>Kniegelenkersatz, n (%)</i>	3 (4,5)	4 (5,9)
<i>Klammern der Wachstumsfugen, n (%)</i>	8 (12,1)	4 (5,9)
<i>Tibiatorsion/Klumpfuß-Korrektur, n (%)</i>	3 (4,5)	3 (4,4)
<u>Frakturen</u>		
<i>Aktive Frakturen, n (%)</i>	8 (12,1)	8 (11,8)
<i>Nicht-aktive Frakturen, n (%)</i>	37 (56,1)	42 (61,8)
<i>Aktive Pseudofrakturen, n (%)</i>	34 (51,5)	29 (42,6)
<i>Nicht-aktive Pseudofrakturen, n (%)</i>	22 (33,3)	24 (35,3)

Studie Gruppe N (FAS)	UX023-CL303	
	Placebo 66	Burosumab 68
<b>Bisherige XLH-Therapie</b>		
<u>Substitutionstherapie erhalten, n (%)</u>	63 (95,4)	62 (91,2)
<i>Substitutionstherapie &lt; 18 Jahre, n (%)</i>	50 (75,7)	50 (73,6)
<i>Substitutionstherapie ≥ 18 Jahre, n (%)</i>	50 (75,7)	48 (70,6)
<u>Dauer der Therapie [Phosphat, Vitamin D] &lt; 18 Jahre</u>		
<i>Phosphat [Jahre], MW (SD)</i>	12,2 (4,38)	12,5 (3,84)
<i>Vitamin D [Jahre], MW (SD)</i>	12,1 (4,69)	13,5 (2,75)
<u>Dauer der Therapie [Phosphat, Vitamin D] ≥ 18 Jahre</u>		
<i>Phosphat [Jahre], MW (SD)</i>	8,2 (9,72)	8,8 (8,52)
<i>Aktives Vitamin D [Jahre], MW (SD)</i>	9,1 (10,79)	9,5 (8,38)
<u>Weitere Therapien</u>		
<i>Analgetika, n (%)</i>	24 (36,4)	32 (47,1)
<i>Entzündungshemmende und antirheumatische Arzneimittel, n (%)</i>	24 (36,4)	28 (41,2)
<i>Vitamine, n (%)</i>	13 (19,7)	20 (29,4)
<i>Psychoanaleptika, n (%)</i>	10 (15,2)	11 (16,2)
<u>Therapien zu Baseline</u>		
<i>Schmerzmittelgebrauch, n (%)</i>	44 (66,7)	47 (69,1)
<i>Opioide, n (%)</i>	13 (19,7)	17 (25,0)
<i>Nichtsteroidale entzündungshemmende Medikamente sowie Acetaminophen/Paracetamol, n (%)</i>	43 (65,2)	47 (69,1)
<i>Schmerzmittel zur Behandlung neuropathischer Schmerzen/Antidepressiva, n (%)</i>	3 (4,5)	4 (5,9)
<i>Andere Schmerzmittel, n (%)</i>	1 (1,5)	7 (10,3)
<b>Stratifizierungscharakteristika (zur Randomisierung)</b>		
<u>Tatsächlicher Stratifizierungsfaktor</u>		
<i>BPI Durchschnittlicher Schmerz, n (%)</i>		
<i>≤ 6,0, MW (SD)</i>	49 (47,2)	48 (70,6)
<i>&gt; 6,0, MW (SD)</i>	17 (25,8)	20 (29,4)
<u>Geplanter Stratifizierungsfaktor</u>		
<i>BPI Schlimmster Schmerz, n (%)</i>		
<i>≤ 6,0, MW (SD)</i>	27 (40,9)	17 (25,0)
<i>&gt; 6,0, MW (SD)</i>	39 (59,1)	51 (75,0)
<u>Region, n (%)</u>		
<i>Nordamerika/EU, n (%)</i>	58 (87,9)	58 (85,3)
<i>Japan, n (%)</i>	5 (7,6)	6 (8,8)
<i>Südkorea, n (%)</i>	3 (4,5)	4 (5,9)

Studie	UX023-CL303	
	Placebo	Burosumab
Gruppe		
N (FAS)	66	68

*MW: Mittelwert, SD: Standardabweichung, FAS: Full Analysis Set, N: Anzahl der Patienten der FAS, n: Anzahl an Studienteilnehmern, (%): prozentualer Anteil von N; TMP/GFR: Tubuläres Maximum der Phosphatrückresorption, BPI: Brief Pain Inventory, BMI: Body-Mass-Index*

a: Bei 5 % der Studienteilnehmer wurde keine PHEX-Mutation zu Baseline festgestellt. Von einem weiteren Studienteilnehmer lagen keine Ergebnisse der Mutationsanalyse vor. Alle Patienten erfüllten jedoch die Einschlusskriterien und wiesen eine FGF23-Konzentration von > 30 pg/ml auf. Es ist darauf hinzuweisen, dass durch Sequenzierungen nicht alle Mutationen im PHEX-Gen identifiziert werden können (z. B. Mutationen in der Promotor-Region, genetische Variationen und kryptische Intron-Spleiß-Mutationen).

*Beschreiben Sie die Studien zusammenfassend. In der Beschreibung der Studien sollten Informationen zur Behandlungsdauer sowie zu geplanter und tatsächlicher Beobachtungsdauer enthalten sein. Sofern sich die Beobachtungsdauer zwischen den relevanten Endpunkten unterscheidet, sind diese unterschiedlichen Beobachtungsdauern endpunktbezogen anzugeben. Beschreiben Sie zudem, ob und aus welchem Anlass verschiedene Datenschnitte durchgeführt wurden oder noch geplant sind. Geben Sie dabei auch an, ob diese Datenschnitte jeweils vorab (d.h. im statistischen Analyseplan) geplant waren. In der Regel ist nur die Darstellung von a priori geplanten oder von Zulassungsbehörden geforderten Datenschnitten erforderlich. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

*Sollte es Unterschiede zwischen den Studien geben, weisen Sie in einem erläuternden Text darauf hin.*

### **Studiendesign und Randomisierung der Studie UX023-CL303**

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde, Placebo-kontrollierte, multizentrische Studie der Phase III. Die Studie wurde an 25 Studienzentren in insgesamt sieben Ländern (acht Studienzentren in den USA, fünf Studienzentren im Vereinigten Königreich Großbritanniens, fünf Studienzentren in Japan, drei Studienzentren in Frankreich, zwei Studienzentren in Südkorea, ein Studienzentrum in Irland und eines in Italien) durchgeführt und schloss Erwachsene mit XLH in einem Alter von 18 – 65 Jahren ein.

Die Studie bestand aus 2 Screening-Visiten, einer Baseline-Visite, einer doppelblinden Placebo-kontrollierten Phase (24 Wochen), einer offenen Behandlungsfortsetzungsphase (24 Wochen) sowie zwei Extensionsphasen (I: 48 Wochen, II: bis zu 53 Wochen) (Abbildung 2).

Teilnehmer, die vor Studienbeginn eine Substitutionstherapie aus oralem Phosphat und aktivem Vitamin D einnahmen, mussten diese nach dem Unterschreiben der Einwilligungserklärung für die Laufzeit der Studie absetzen. Bei diesen Studienteilnehmern fand die zweite Screening-Visite frühestens 14 Tage nach Beendigung der Therapie statt. Die Baseline-Visite konnte entweder am Tag nach der zweiten Screening-Visite erfolgen oder bis zu 7 Tage später.

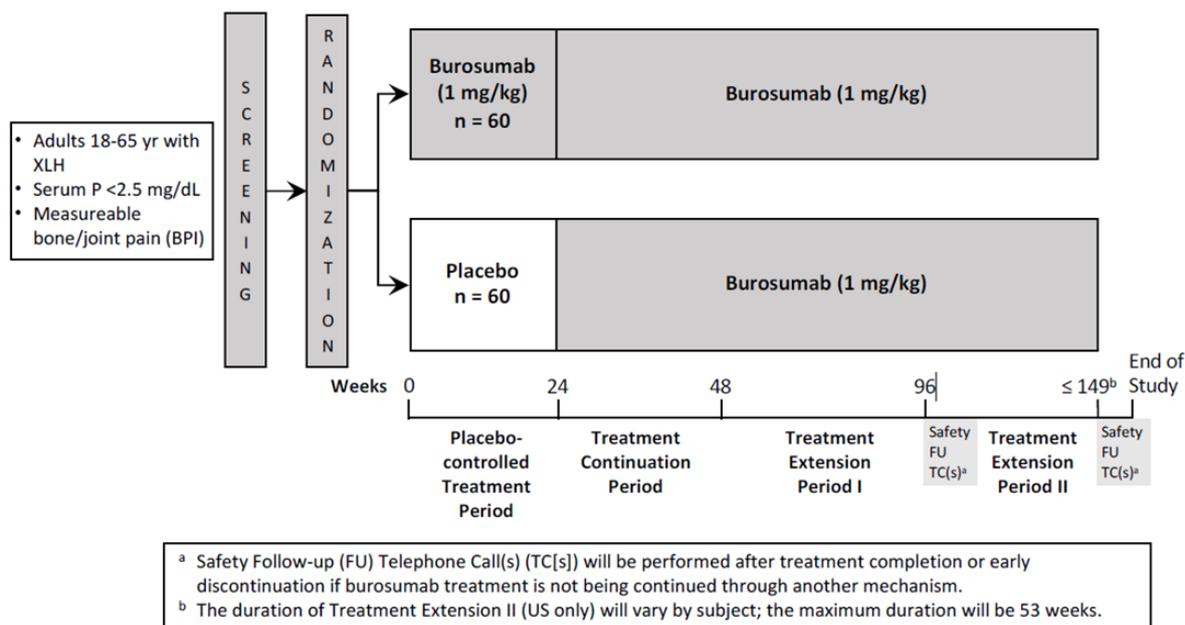


Abbildung 2: Studiendesign der Studie UX023-CL303

Ursprünglich sollte die Randomisierung in der Studie UX023-CL303 anhand der Frage 3 *BPI (Brief Pain Inventory) Schlimmster Schmerz* des patientenberichteten Fragebogens *BPI* und der Region stratifiziert werden. Hierfür wurde anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite so wie am Tag der Visite der Schmerz anhand einer verkürzten Form des Fragebogens *BPI* bestimmt. Die verkürzte Form des Fragebogens enthielt Frage 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-22). Das für die Stratifizierung verwendete *Interactive Web Response System (IWRS)* war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz* ≤ 6,0 oder > 6,0) sondern Frage 3 (*BPI Durchschnittlicher Schmerz* ≤ 6,0 oder > 6,0) verwendet wurde.

Tabelle 4-22: Nummerierung der einzelnen Items des *BPI*-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	_b	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	_b	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.
9 A – G	_b	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:
a: IWRS: <i>Interactive web response system</i>		
b: für die Stratifizierung wurde eine verkürzte Form des <i>BPI</i> -Fragebogens verwendet, der nur die Items der Domäne Schmerzintensität enthielt.		

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Um die Robustheit der Ergebnisse zu überprüfen, wurden zusätzlich Sensitivitätsanalysen für den primären Endpunkt *Serumphosphat* (mg/dl) und die drei sekundären Endpunkte *BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion* durchgeführt (siehe Abschnitt 4.2.5.4). Diese Analysen waren den primären Analysen sehr ähnlich und bestätigen somit, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse in der Studie hat. Der Stratifizierungsfaktor Region wurde aufgrund von operativen und logistischen Gründen zusätzlich zu dem Merkmal Schmerz mit aufgenommen. Da lediglich eine geringe Anzahl an Studienteilnehmern aus Japan und Korea zu erwarten war, sollte dies eine Gleichverteilung auf die beiden Behandlungsgruppen Burosumab und Placebo gewährleisten. Details zur Randomisierung sind in der Randomisierungsspezifikation vom 27.07.2015 zu finden [52].

## Patientencharakteristika der Studienpopulation

### *Baseline-Charakteristika und krankheitsspezifische Folgeerkrankungen von XLH*

Die Zeit seit dem ersten Auftreten von XLH-assoziierten Symptomen betrug in der Studienpopulation durchschnittlich 37,2 Jahre (durchschnittliches Alter der Studienteilnehmer zu Studienbeginn: 39,99 Jahre, Alter bei dem Auftreten erster Symptome: ca. 3 Jahre), wobei

die Diagnose, aufgrund des komplexen Krankheitsbildes, erst später erfolgte (vor 31,41 Jahre in einem Alter von ca. 9 Jahren). 93,3 % der Studienteilnehmer erhielt die für Kinder empfohlene Substitutionstherapie (orales Phosphat) [3]. 74,6 % der Erwachsenen wurden bereits vor Abschluss des Skelettwachstums mit der Substitutionstherapie behandelt und 73,1 % im Erwachsenenalter. Die durchschnittliche Einnahme von Phosphat betrug 16,5 Jahre und von aktivem Vitamin D 18,2 Jahre. Trotz dieser Therapie wiesen alle Studienteilnehmer zu Beginn der Studie schwere Folgeerkrankungen der XLH auf. Hierzu zählen Enthesopathien (99,3 %), Skelettdeformationen (94,0 %), Arthrose (63,4 %) sowie Verkalkungen der Wirbelsäule (62,7 %). Die ausgeprägte Osteomalazie wurde zudem an dem häufigen Vorhandensein von nicht-aktiven Frakturen (59 %), nicht-aktiven Pseudofrakturen (34,3 %), aktiven Frakturen (11,9 %) und aktiven Pseudofrakturen (47,0 %) deutlich. Darüber hinaus lag die *Serumphosphat-Konzentration* bei allen Studienteilnehmern (100 %) vor Studienbeginn unterhalb (durchschnittlich 1,98 mg/dl) der unteren Grenze des Normbereichs von Erwachsenen (LLN:  $\geq 2,5$  mg/dl). Die tubuläre Phosphat-Rückresorption (*Tmp/GFR*) betrug 1,639 mg/dl (Normbereich: 2,5 – 4 mg/dl) und bestätigt ebenfalls wie die niedrige Serumkonzentration von *1,25(OH)<sub>2</sub>D* (33 pg/ml, Normbereich: 18 – 66 pg/ml) die starke Ausprägung der XLH-Erkrankung. Der *Brief Pain Inventory (BPI) Durchschnittliche Schmerz*, gemessen anhand des patientenberichteten Fragebogens *BPI* über 7 Tage vor Studienbeginn, wies auf einer Skala von 0 – 10 einen Wert von 5,1 auf. Dagegen wurde der *BPI Schlimmste Schmerz* innerhalb der letzten 24 Stunden als 6,68 klassifiziert und zeigte eine starke Schmerzbelastung der Erwachsenen mit XLH auf.

### ***Folgeerkrankungen durch die jahrelange Einnahme der Substitutionstherapie***

23,1 % (31/134) der Studienteilnehmer berichteten bereits an Erkrankungen der Niere gelitten zu haben. Dabei handelte es sich in 51,6 % (16/31) der Fälle um eine Nephrokalzinose und in 58,06 % (18/31) der Fälle um Nephrolithiasis. 7 (5,22 %) Studienteilnehmer hatten zudem bereits an einem Hyperparathyreoidismus gelitten. Dabei handelte es sich bei 3 der 7 Studienteilnehmer um einen sekundären Hyperparathyreoidismus.

### ***Krankheitsbefund von XLH zu Studienbeginn***

Der Krankheitsbefund beinhaltete typische Krankheitsmerkmale von XLH wie z. B. Knochen- und Skelettdeformationen, Knochen- und Gelenkschmerzen sowie ein erhöhtes Level an FGF23 ( $> 30$  pg/ml, 100 % der Studienteilnehmer) oder das Vorhandensein einer *PHEX*-Mutation (95 %).

Die Patienten der Studie UX023-CL303 waren hinsichtlich der Baseline-Charakteristika sehr homogen (Tabelle 4-21). Die beiden Behandlungsgruppen, die entweder eine Burosumab- oder Placebo-Gabe erhielten, sind somit vergleichbar. Dies wurde zusätzlich durch die Stratifizierung nach *BPI Durchschnittlicher Schmerz* (Abschnitt 4.2.5.2) und Region sichergestellt.

## Behandlung

### **Intervention: Burosumab**

Während der 24-wöchigen doppelblinden, Placebo-kontrollierten Phase erhielten die Studienteilnehmer der Burosumab-Gruppe alle 4 Wochen 1 mg/kg KG Burosumab subkutan. Falls eine Serumphosphat-Konzentration von  $> 4,5$  mg/dl überschritten wurde, wurde der Prüfarzt entblindet, aber nicht die Studienteilnehmer selbst, sodass die Burosumab-Dosis angepasst werden konnte.

### **Vergleichstherapie: Placebo (= best supportive care)**

Die Placebo-Gruppe erhielt stattdessen alle 4 Wochen subkutan Placebo, das sich aus den gleichen Bestandteilen wie die Prüfmedikation zusammensetzte, jedoch nicht die aktive Prüfsubstanz enthielt.

### **Implementierung der zweckmäßigen Vergleichstherapie (zVT) in der Studie UX023-CL303**

Die zweckmäßige Vergleichstherapie (zVT) für das Anwendungsgebiet XLH wurde in einem Beratungsgespräch am 11.03.2020 beim G-BA erfragt. Der G-BA gab in diesem Gespräch als zVT eine Phosphatsubstitution mit aktivem Vitamin D an [7]. Die Kyowa Kirin GmbH folgt diesen Angaben jedoch nicht, da nach eigener Auffassung bis zum heutigen Zeitpunkt keine Leitlinien oder einheitlichen Therapieempfehlungen zur Behandlung von Erwachsenen mit XLH vorliegen. Weiterhin weisen, trotz mehrjähriger Vorbehandlung mit Phosphat, die meisten Studienteilnehmer schwere Folgeerkrankungen auf. Aufgrund der mangelnden Wirksamkeit stellt die Phosphatsubstitution eine unzureichende Therapie dar, deren Fortsetzen nicht einer adäquaten Umsetzung der zVT entsprechen kann. Da keine den Pathomechanismus adressierende Therapieoption existiert, ist die Kyowa Kirin GmbH der Meinung, dass lediglich „best supportive care“ die alleinige zVT darstellt. Als „best supportive care“ wird die Therapie verstanden, die eine bestmögliche, patientenindividuell optimierte unterstützende Behandlung zur Linderung von Symptomen und Verbesserung der Lebensqualität gewährleistet. Eine ausführliche Begründung ist in Modul 3 Abschnitt 3.1.2 aufgeführt.

Die Folgeerkrankungen der XLH wie z. B. Skelettdeformationen, Osteomalazie, Arthrose, Enthesopathien, Spinalkanalstenosen und (Pseudo-)Frakturen sind mit starken Schmerzen und Steifheit assoziiert und führen zu erheblichen Einschränkungen in der Mobilität, der körperlichen Funktionsfähigkeit und der Lebensqualität. Zusätzlich weisen die XLH-Patienten eine ausgeprägte Fatigue auf. In einer Umfrage zur Krankheitslast bei der XLH gaben 67 % der Befragten an, mindestens einmal pro Woche Schmerzmittel einzunehmen, um die Schmerzen zu kompensieren [53]. Dabei handelte es sich in 69 % der Fälle um die Einnahme von rezeptfreien Schmerzmitteln (nichtsteroidale entzündungshemmende Arzneimittel, Naproxen, Paracetamol) und in 21 % der Fälle um verschreibungspflichtige Schmerzmittel einschließlich von Opioiden (Oxycodon, Oxycodon in Kombination mit Acetaminophen, Acetaminophen in Kombination mit Codein, Hydrocodon Bitartrat, Kortison, Hydrocodon-Bitartrat-Acetaminophen) [53]. Zusätzlich erforderte das hohe Ausmaß an Schmerzen bei einem Teil der Befragten die regelmäßige Einnahme von Antidepressiva. Die schlechte Zahngesundheit, die sich u. a. in Form von Zahnabszessen und exzessiven Karies äußert, bedingt darüber hinaus den

Einsatz von Antibiotika. Neben den skelettalen und rheumatischen Beschwerden weisen XLH-Patienten außerdem ein erhöhtes Risiko für kardiovaskuläre Erkrankungen wie z. B. Bluthochdruck und linksventrikuläre Hypertrophie auf [54]. Des Weiteren ist die pathologische FGF23-Konzentration mit Übergewicht und Fettleibigkeit assoziiert [55]. Hinzu kommen die Folgeerkrankungen einer jahrelangen Phosphatsubstitution, die sich in Form von Nephrokalzinosen, Nephrolithiasis und Hyperparathyreoidismus äußern. Die XLH-Patienten leiden somit an einem breiten Spektrum von Symptomen und Folgeerkrankungen, die eine multidisziplinäre Behandlung erforderlich machen.

Die Umsetzung von „*best supportive care*“ wurde in der Studie UX023-CL303 durch die zusätzliche Gabe von Begleittherapien bei Bedarf gewährleistet. Diese umfassen u. a.:

- Nichtsteroidale entzündungshemmende Medikamente
- Opioide
- Narkotische Schmerzmittel

Ein Auszug aus der Begleitmedikation in der Studie UX023-CL303 ist in Tabelle 4-23 und Tabelle 4-24 aufgeführt. Eine detaillierte Auflistung ist im CSR der Studie UX023-CL303 zu finden [5].

Der Gebrauch von Schmerzmitteln (verschreibungspflichtige und rezeptfreie Schmerzmittel) wurde in einem Schmerzmittel-Tagebuch in einem Zeitraum von sieben Tagen vor Baseline bzw. vor der Studienvisite zu Woche 12, 24, 36 und 148 dokumentiert. Die Dosis an Schmerzmitteln durfte in der Studie nicht erhöht werden und war auf 60 mg Morphinäquivalente / Tag beschränkt.

Begleitmedikationen und -therapien wurden bei jedem Studienbesuch im Prüfzentrum und gegebenenfalls bei Hausbesuchen im eCRF des Studienteilnehmers überprüft und aufgezeichnet. Die Studienteilnehmer mussten sich zudem bereit erklären, die Dosierung und das Behandlungsschema der chronischen Schmerzmedikation während des gesamten Placebo-kontrollierten Behandlungszeitraums der Studie (24 Wochen) konstant zu halten. Zudem wurden alle innerhalb eines Zeitraums von 30 Tagen vor dem Screening eingenommenen Schmerzmedikamente, einschließlich der im Schmerztagbuch der Probanden erfassten Medikamente als Begleitmedikation erfasst.

Tabelle 4-23: Auszug aus den Begleitmedikation während der doppelblinden Phase der Studie UX023-CL303

Begleitmedikation	Burosumab n (%)	Placebo n (%)
<b>Personen mit mindestens einer Begleitmedikation</b>	<b>67 (98,5)</b>	<b>62 (93,9)</b>
Analgetika	53 (77,9)	45 (68,2)
Entzündungshemmende und antirheumatische Arzneimittel	49 (72,1)	40 (60,6)
Antibiotika	25 (36,8)	30 (45,5)

<b>Begleitmedikation</b>	<b>Burosumab n (%)</b>	<b>Placebo n (%)</b>
Vitamine	29 (42,6)	21 (31,8)
Mittel bei säurebedingten Erkrankungen	17 (25,0)	18 (27,3)
Psychoanaleptika	14 (20,6)	16 (24,2)
Renin-Angiotensin-System beeinflussende Arzneimittel	12 (17,6)	10 (15,2)
Kortikosteroide für den systemischen Gebrauch	11 (16,2)	7 (10,6)
Muskelrelaxantien	10 (14,7)	7 (10,6)
Beta-Blocker	6 (8,8)	9 (13,6)
Calcium-Kanalblocker	5 (7,4)	6 (9,1)
Topische Produkte für Gelenk- und Muskelschmerzen	6 (8,8)	5 (7,6)
Antidiabetika	2 (2,9)	3 (4,5)
Arzneimittel bei Adipositas	3 (4,4)	1 (1,5)
Antiepileptika	2 (2,9)	2 (3,0)

Tabelle 4-24: Schmerzmittelgebrauch bei Studienvsiten

	<b>Burosumab n (%)</b>	<b>Placebo n (%)</b>
<b>Jeglicher Schmerzmittelgebrauch zu Baseline</b>	<b>47 (69,1)</b>	<b>44 (66,7)</b>
Opioide	17 (25,0)	13 (19,7)
Nicht-Opioide Schmerzmittel	47 (69,1)	43 (65,2)
Neuropathische Schmerzmedikamente/Antidepressiva	4 (5,9)	3 (4,5)
Andere Schmerzmittel	7 (10,3)	1 (1,5)

Die Begleittherapie in der Studie UX023-CL303 setzt sich somit überwiegend aus Schmerzmitteln und entzündungshemmenden sowie antirheumatischen Arzneimitteln zusammen, um die Symptome der schwerwiegenden Folgeerkrankungen der XLH zu lindern und entspricht der gängigen klinischen Praxis bei der Behandlung von XLH-Patienten [19, 53, 56, 57].

Nach Auffassung der Kyowa Kirin GmbH stellt Placebo in Kombination mit einer Begleittherapie, wie in der Studie UX023-CL303 eingesetzt, die geeignete Umsetzung der zVT „best supportive care“ dar.

### ***Behandlungsfortsetzungsphase und Extensionsphase***

Während der Behandlungsfortsetzungsphase und den beiden Extensionsphasen erhielten alle Studienteilnehmer Burosumab, und die Verblindung wurde aufgehoben. Die Gruppenzuteilung während der 24-wöchigen doppelblinden Phase wurde jedoch erst nach Beendigung der Behandlungsfortsetzungsphase dem Studienpersonal, Prüfarzt und Studienteilnehmern

mitgeteilt. Lediglich eine Auswahl an Studienpersonal wurde direkt nach der 24-wöchigen doppelblinden Phase entblindet, um die Analysen durchführen zu können.

## **Expositionszeiten**

### ***Behandlungsphase (Woche 0 – 24)***

#### *Burosumab*

Alle Erwachsenen der Burosumab-Gruppe erhielten die 24-wöchige im Studienprotokoll spezifizierte Behandlung mit Burosumab und wurden vollständig in die Analysen einbezogen. Die durchschnittliche Expositionszeit mit Burosumab lag bei 168,1 Tage. Bezogen auf die Gesamtzeit der Exposition aller Studienteilnehmer während der doppelblinden Phase der Studie, ist dies gleichzusetzen mit einer Expositionszeit von 31,30 Jahren. Die mediane Dosis an Burosumab betrug in dieser Zeit 1,0 mg/kg KG.

#### *Placebo*

Alle Erwachsenen der Placebo-Gruppe erhielten die 24-wöchige im Studienprotokoll spezifizierte Behandlung mit Placebo und wurden vollständig in die Analysen einbezogen. Die Expositionszeit gegenüber der Prüfsubstanz, lag ähnlich wie bei der Burosumab-Gruppe, durchschnittlich bei 167,9 Tage. Die Gesamtzeit der Exposition, bezogen auf alle Studienteilnehmer, betrug während der doppelblinden Phase 30,34 Jahre.

#### *Einhaltung der im CSR vorgeschriebenen Behandlung*

Burosumab bzw. Placebo wurde durch geschultes Personal entweder bei Hausbesuchen oder Klinikvisiten verabreicht. Bei 7,3 % (n = 5) der Studienteilnehmer der Burosumab-Gruppe wurde innerhalb der 24-wöchigen Studienphase der Prüfarzt aufgrund einer erhöhten *Serumphosphat-Konzentration* entblindet (mind. 4,5 mg/dl). Bei 4 der 5 Studienteilnehmer wurde die Burosumab-Dosis daraufhin auf die Hälfte reduziert, bei einem Studienteilnehmer war dagegen zusätzlich eine weitere Dosisreduktionen notwendig. Darüber hinaus verpassten in der Placebo-Gruppe zwei Studienteilnehmer die Verabreichung der Prüfsubstanz.

## **Übertragbarkeit auf den Versorgungskontext**

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern wird die doppelblinde Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zudem handelt es sich bei der XLH um eine genetisch bedingte Erkrankung, welche unabhängig von äußeren Einflüssen entsteht, sodass keine Anzeichen vorliegen, die gegen eine Übertragbarkeit sprechen [58]. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt. Die

Übertragbarkeit der Ergebnisse der Studie UX023-CL303 auf den deutschen Versorgungskontext wurde zudem bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 durch den G-BA bestätigt [1, 46].

#### 4.3.1.2.2 Verzerrungspotenzial auf Studienebene

Bewerten Sie das Verzerrungspotenzial der RCT auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-25: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial auf Studienebene
			Patient	Behandelnde Personen			
UX023-CL303	ja	ja	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

#### UX023-CL303

##### **Randomisierung**

Bei der Studie UX023-CL303 handelt es sich um eine doppelblinde, Placebo-kontrollierte, multizentrische Studie der Phase III, die Erwachsene mit XLH im Alter von 18 – 65 Jahre einschloss. Die Randomisierung erfolgte 1:1 in die beiden Behandlungsgruppen (Burosumab und Placebo) und wurde anhand des *Interactive Web Response System* (IWRS) durchgeführt (Details zur Randomisierung sind in der Randomisierungsspezifikation vom 27.07.2015 zu finden [52]). Die Stratifizierung sollte ursprünglich anhand der Frage 3 *Brief Pain Inventory* (BPI) *Schlimmster Schmerz* des patientenberichteten Fragebogens BPI und der Region erfolgen. Der Stratifizierungsfaktor Region wurde aufgrund von operativen und logistischen Gründen zusätzlich zu dem Merkmal Schmerz aufgenommen. Da lediglich eine geringe Anzahl an Studienteilnehmern aus Japan und Korea zu erwarten war, sollte dies eine Gleichverteilung auf die beiden Behandlungsgruppen Burosumab und Placebo gewährleisten. Der Stratifizierungsfaktor Schmerz wurde anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens BPI bestimmt. Die verkürzte Form des Fragebogens enthielt Frage 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als

Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-26). Das für die Stratifizierung verwendete *Interactive web response system* (IWRS) war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*) verwendet wurde (Anhang 4-E 8b).

Tabelle 4-26: Nummerierung der einzelnen Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	- <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	- <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.
9 A – G	- <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:

a: IWRS: *Interactive web response system*  
b: für die Stratifizierung wurde eine verkürzte Form des BPI-Fragebogens verwendet, der nur die Items der Domäne Schmerzintensität enthielt.

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Da auch die Sensitivitätsanalysen, die durchgeführt wurden, um den Einfluss der fehlerhaften Anpassung

des IWRS zu prüfen, die Robustheit der Ergebnisse bestätigen, wird insgesamt von einer adäquaten Erzeugung der Randomisierungssequenz ausgegangen.

### **Verblindung**

Sowohl den Studienteilnehmern, dem Studienpersonal als auch dem Prüfarzt war die Gruppenzuteilung während der 24-wöchigen, doppelblinden Phase der Studie nicht bekannt. Beide Prüfsubstanzen, Placebo und Burosumab, sahen darüber hinaus gleich aus. Die patientenberichteten Fragebögen wurden von den Studienteilnehmern selbst ausgefüllt, und biochemische Analysen erfolgten objektiv und verblindet in einem zentralen Labor. Somit ist eine ergebnisunabhängige Berichterstattung über den gesamten Zeitraum der doppelblinden Phase der Studie gewährleistet.

### **Sonstige Aspekte**

Für die Analysen aller Endpunkte wurde die FAS-Population eingesetzt. Hierbei handelt es sich um die randomisierte Studienpopulation, die während der doppelblinden Phase mindestens eine Gabe der Prüfsubstanz erhielt. Fehlende Werte wurden bei der Analyse als fehlend betrachtet. Um die Robustheit der Ergebnisse trotz fehlerhafter Stratifizierung zu Studienbeginn oder fehlender Werte zu überprüfen, wurden Sensitivitätsanalysen für den primären Endpunkt *Serumphosphat (mg/dl)* und die drei sekundären Wirksamkeitsendpunkte *Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Steifheit*, *WOMAC Physische Funktion* und *BPI Schlimmster Schmerz* durchgeführt.

#### *Sensitivitätsanalysen aufgrund der fehlerhaften Stratifizierung zu Studienbeginn*

Um einen möglichen Einfluss der fehlerhaften Stratifizierung auf den primären Wirksamkeitsendpunkt *Serumphosphat Prozentsatz erreichter Normwert* und die sekundären Wirksamkeitsendpunkte *WOMAC Steifheit* und *WOMAC Physische Funktion* zu analysieren, wurde eine Sensitivitätsanalyse durchgeführt, die für den ursprünglich geplanten Stratifizierungsfaktor *BPI Schlimmster Schmerz* adjustierte. Für den sekundären Wirksamkeitsendpunkt *BPI Schlimmster Schmerz* adjustierte das *Generalized estimation equation (GEE)*-Modell dagegen für den *BPI Durchschnittlichen Schmerz*. Diese Analysen bestätigen, dass die fehlerhafte Stratifizierung keinen relevanten Einfluss auf die Ergebnisse der Studie UX023-CL303 hat.

#### *Sensitivitätsanalysen aufgrund fehlender Post-Baseline-Werte*

Hierbei wurde das Verfahren des *last observation carried forward (LOCF)*, *baseline observation carried forward (BOCF)* und *modified baseline observation carried forward (mBOCF)* eingesetzt. Die Ergebnisse waren übereinstimmend mit der Primäranalyse und bestätigen, dass fehlende Werte keinen relevanten Einfluss auf die Analysen in der Studie haben. Somit erfolgte während der gesamten Studiendauer eine adäquate Umsetzung des ITT-Prinzips.

Es wurden keine sonstigen endpunktübergreifenden Aspekte identifiziert, die zu einer Verzerrung der Studienergebnisse geführt haben könnten. Zusammenfassend ist somit bei der Studie UX023-CL303 von einem **niedrigen Verzerrungspotenzial** auszugehen.

#### 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens beruht. Geben Sie dabei an, welche dieser Endpunkte in den relevanten Studien jeweils untersucht wurden. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-27: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

Endpunktkategorie	Endpunkt	UX023-CL303
Mortalität	Im Rahmen der Sicherheit erhoben	_ <sup>a</sup>
Morbidity	Serumphosphat	X
	Gehfähigkeit: <i>6-Minuten-Gehdistanz Test (6MWT)</i>	X
	Steifheit und physische Funktion: <i>Western Ontario and McMaster Universities Osteoarthritis (WOMAC)</i>	X
	Schmerzerhebung: <i>Brief Pain Inventory (BPI)</i>	X
	Fatigue: <i>Brief Fatigue Inventory</i>	X
	Erfassung von Therapieeffekten: <i>Patient Global Impression of Improvement (PGI-I)</i>	X
Gesundheitsbezogene Lebensqualität	_ <sup>b</sup>	_ <sup>b</sup>
Sicherheit	Unerwünschte Ereignisse, schwerwiegende unerwünschte Ereignisse, Therapieabbruch wegen unerwünschter Ereignisse und Todesfälle sowie unerwünschte Ereignisse von besonderem Interesse	X
X: Endpunkt wurde in Studie UX023-CL303 erhoben -: Endpunkt wurde nicht erhoben a: Todesfälle wurden im Rahmen der unerwünschten Ereignisse erhoben. b: Die Gesundheitsbezogene Lebensqualität wurde in der Studie UX023-CL303 nicht erhoben.		

#### 4.3.1.3.1 Ergebnisse patientenrelevanter Endpunkte

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Meta-Analyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind inkl. Angaben zur Häufigkeit von und zum Umgang mit nicht oder nicht vollständig beobachteten Patienten (bei Verlaufsbeobachtungen pro Messzeitpunkt)
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde.

Unterschiedliche Beobachtungszeiten zwischen den Behandlungsgruppen sollen durch adäquate Analysen (z.B. Überlebenszeitanalysen) adressiert werden, und zwar für alle Endpunkte (einschließlich UE nach den nachfolgend genannten Kriterien), für die eine solche Analyse aufgrund deutlich unterschiedlicher Beobachtungszeiten erforderlich ist.

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden. Dabei ist für jeden Endpunkt, für den eine solche Analyse durchgeführt wird, eine separate Kaplan-Meier-Kurve darzustellen.

Zu mit Skalen erhobenen patientenberichteten Endpunkten (z.B. zur gesundheitsbezogenen Lebensqualität oder zu Symptomen) sind immer auch die Werte im Studienverlauf anzugeben, auch als grafische Darstellung, sowie eine Auswertung, die die über den Studienverlauf ermittelten Informationen vollständig berücksichtigt (z.B. als Symptomlast über die Zeit, geschätzt mittels MMRM-Analyse [falls aufgrund der Datenlage geeignet]).

Zu unerwünschten Ereignissen (UE) sind folgende Auswertungen vorzulegen:

1. Gesamtrate UE,
2. Gesamtrate schwerwiegender UE (SUE),

3. Gesamtrate der Abbrüche wegen UE,

4. Gesamtraten von UE differenziert nach Schweregrad, sofern dies in der/den relevante/n Studie/n erhoben wurde (z.B. gemäß CTCAE und/oder einer anderen etablierten bzw. validierten indikationsspezifischen Klassifikation) einschließlich einer Abgrenzung schwerer und nicht schwerer UE,

5. zu den unter 1, 2 und 4 genannten Kategorien (UE ohne weitere Differenzierung, SUE, UE differenziert nach Schweregrad) soll zusätzlich zu den Gesamtraten die Darstellung nach Organsystemen und Einzelereignissen (als System Organ Class [SOCs] und Preferred Terms [PT] nach MedDRA) jeweils nach folgenden Kriterien erfolgen:

- UE (unabhängig vom Schweregrad): Ereignisse, die bei mindestens 10% der Patienten in einem Studienarm aufgetreten sind

- Schwere UE (z.B. CTCAE-Grad  $\geq 3$ ) und SUE: Ereignisse, die bei mindestens 5% der Patienten in einem Studienarm aufgetreten sind

- zusätzlich für alle Ereignisse unabhängig vom Schweregrad: Ereignisse, die bei mindestens 10 Patienten UND bei mindestens 1 % der Patienten in einem Studienarm aufgetreten sind.

6. A priori definierte UE von besonderem Interesse [AESI]) sowie prädefinierte SOC-übergreifende UE-Auswertungen (z.B. als Standardised MedDRA Queries, SMQs) sollen unabhängig von der Ereignisrate dargestellt werden und zwar differenziert nach Schweregrad (dargestellt als Gesamtrate und differenziert nach Schweregrad, nicht schwer, schwer, schwerwiegend).

7. zu Kategorie 3: Die Abbruchgründe auf SOC/PT-Ebene müssen vollständig, jedoch nur deskriptiv dargestellt werden.

Sofern bei der Erhebung unerwünschter Ereignisse erkrankungsbezogenen Ereignisse (z. B. Progression, Exazerbation) berücksichtigt werden (diese Ereignisse also in die UE-Erhebung eingehen), sollen für die Gesamtraten (UE, schwere UE und SUE) zusätzliche UE-Analysen durchgeführt werden, bei denen diese Ereignisse unberücksichtigt bleiben. Alle Auswertungen zu UE können auch in einem separaten Anhang des vorliegenden Modul 4 dargestellt werden. Dabei kann die Ausgabe der Statistik-Software unverändert verwendet werden, sofern diese alle notwendigen Angaben enthält. Eine Darstellung ausschließlich in Modul 5 ist nicht ausreichend. Davon unbenommen sind die Gesamtraten (UE, schwere UE, SUE und Abbrüche wegen UE), sowie die für die Gesamtaussage zum Zusatznutzen herangezogenen Ergebnisse im vorliegenden Abschnitt darzustellen.

Auswertungen zu den im Abschnitt 4.3.1.2.1 aufgeführten Datenschnitten sollen vollständig, d.h. für alle erhobenen relevanten Endpunkte, durchgeführt und vorgelegt werden. Das gilt auch dann wenn ein Datenschnitt ursprünglich nur zur Auswertung einzelner Endpunkte geplant war. Auf die Darstellung der Ergebnisse einzelner Endpunkte eines Datenschnitts bzw. eines gesamten Datenschnitts kann verzichtet werden, wenn hierdurch kein wesentlicher

Informationsgewinn gegenüber einem anderen Datenschnitt zu erwarten ist (z. B. wenn die Nachbeobachtung zu einem Endpunkt bereits zum vorhergehenden Datenschnitt nahezu vollständig war oder ein Datenschnitt in unmittelbarer zeitlicher Nähe zu einem anderen Datenschnitt liegt).

Falls für die Auswertung eine andere Population als die ITT-Population herangezogen wird, soll diese benannt (z.B. Safety-Population) und definiert werden.

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

*Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.*

#### 4.3.1.3.1.1 Serumphosphat – RCT

Tabelle 4-28: Operationalisierung von *Serumphosphat*

Studie	Operationalisierung
UX023-CL303	<p>Zur Bestimmung der <i>Serumphosphat-Konzentration in mg/dl</i> wurden die Blutproben nach einer Fastenzeit von 8 h über Nacht sowohl zu Studienvsitate 1 und 2, Baseline und Post-Baseline-Zeitpunkten entnommen. Aufgrund der fluktuierenden Serumphosphat-Konzentrationen, die mit dem circadianen Rhythmus assoziiert sind, wurden die Blutproben morgens möglichst jeweils zur gleichen Uhrzeit entnommen. Die Auswertung erfolgte verblindet durch ein zentrales Labor.</p> <p>Die <i>Serumphosphat-Konzentration in mg/dl</i> wurde während der doppelblinden Phase der Studie sowohl zu Baseline als auch zu Woche 1, 2, 6, 10, 14, 18, 22 und 24 gemessen.</p> <p>Zusätzlich zu der <i>Serumphosphat-Konzentration in mg/dl</i> wurde der prozentuale Anteil an Studienteilnehmern bestimmt, die, gemittelt über die Dosiszyklen zwischen Baseline und Woche 24, eine <i>Serumphosphat-Konzentration von <math>\geq 2,5</math> mg/dl</i> in der Mitte des Dosiszyklus erreichten.</p> <p><u>Analysepopulation:</u> FAS</p> <p><u>Statistisches Modell:</u></p> <ul style="list-style-type: none"> <li>• Auswertungen zur mittleren Änderung der <i>Serumphosphat-Konzentration</i> wurden zu Woche 12 und Woche 24 mit einem GEE-Modell bestimmt. Bei dem GEE-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie die Baseline <i>Serumphosphat-Konzentration</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline in der <i>Serumphosphat-Konzentration</i>. Als Kovarianzstruktur wurde eine <i>Compound Symmetry Matrix</i> verwendet.</li> <li>• Responderanalyse Anteil an Studienteilnehmern mit einer <i>Serumphosphat-Konzentration <math>\geq 2,5</math> mg/dl</i>: Prospektiv geplant und im Studienbericht der Studie UX023-CL303 dargestellt ist die Auswertung des <i>Odds Ratio</i> (OR) mittels Cochran-Mantel-Haenszel-(CMH-)Test. Für das Dossier wurde <i>post-hoc</i> das OR inklusive 95% KI und p-Wert anhand eines logistischen Regressionsmodells bestimmt (Woche 24). Die Behandlungsgruppe, der tatsächliche Stratifizierungsfaktor (<i>BPI durchschnittlicher Schmerz</i>) und die Altersgruppen wurden als unabhängige Variablen in das Modell aufgenommen. Das adjustierte RR (Adjustierung analog zum OR) und die dazugehörigen Konfidenzintervalle wurden <i>post-hoc</i> nach der Methode von Zhang und Yu bestimmt (adjustiert für die Behandlungsgruppe, den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) und die Altersgruppen (siehe Abschnitt 4.2.5.2) [45]. Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation und der p-Wert mittels Fishers exaktem Test berechnet.</li> </ul> <p><u>Darstellung im Dossier:</u></p> <ul style="list-style-type: none"> <li>• Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen</li> <li>• Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier anhand von Hedges'g geprüft</li> <li>• Responder-Analyse: Anteil der XLH-Patienten, die eine <i>Serumphosphat-Konzentration</i> im unteren Normbereich (<math>\geq 2,5</math> mg/dl) zu Woche 24 erlangten</li> <li>• Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn: Anteil der XLH-Patienten, die eine <i>Serumphosphat-Konzentration</i> im unteren</li> </ul>

Studie	Operationalisierung
	Normbereich ( $\geq 2,5$ mg/dl) zu Woche 24 erlangten adjustiert nach dem ursprünglichen Stratifizierungsfaktor <i>BPI Schlimmster Schmerz</i>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-29: Bewertung des Verzerrungspotenzials für *Serumphosphat* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde Placebo-kontrollierte, multizentrische Studie der Phase III, die Erwachsene im Alter von 18 – 65 Jahren einschließt. Während der doppelblinden 24-wöchigen Phase waren die Studienteilnehmer, das Studienpersonal und die Prüfarzte gegenüber der Behandlung verblindet. Placebo und Burosumab wurden jeweils im gleichen Zeitintervall subkutan appliziert und unterschieden sich in ihrer Zusammensetzung bis auf die Prüfsubstanz nicht. Die Blutproben zur Messung des primären Endpunkts *Serumphosphat* wurden nach einer 8-stündigen Fastenzeit über Nacht sowie möglichst immer zur gleichen Uhrzeit entnommen. Dies sollte das Auftreten von fluktuierenden Konzentrationen vermeiden, die mit dem circadianen Rhythmus assoziiert sind. Die Auswertung erfolgte verblindet durch ein zentrales Labor. Da alle Studienteilnehmer während der 24-wöchigen Phase mindestens einmal die Prüfsubstanz erhielten, umfasst die FAS die gesamte Studienpopulation. Somit fand eine ergebnisunabhängige Berichterstattung unter adäquater Umsetzung des ITT-Prinzips statt.

Aufgrund der zu Beginn fehlerhaften durchgeführten Stratifizierung (Anhang 4-F) wurden für den primären Endpunkt *Serumphosphat* Sensitivitätsanalysen durchgeführt, um die Robustheit der Ergebnisse zu überprüfen. Diese Analysen waren jedoch übereinstimmend mit den Primäranalysen und bestätigen daher, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse der Studie UX023-CL303 hat.

Zusammenfassend liegt für den Endpunkt *Serumphosphat* ein **niedriges Verzerrungspotenzial** vor.

*Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend*

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-30: Ergebnisse für *Serumphosphat* aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	<i>Serumphosphat</i> (mg/dl)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b>Gesamtrate</b>								
Placebo (N=66)	66 1,92 (0,316)	64 1,91 (0,312)	64 2,06 (0,356)	66 1,92 (0,316)	66 2,07 (0,343)	64 0,14 (0,037) 0,01 (0,070)	66 0,15 (0,043) 0,03 (0,070)	0,40 [0,28; 0,51] < 0,0001 H'g = 0,922 [0,566; 1,278]
Burosumab (N=68)	68 2,03 (0,304)	67 2,03 (0,306)	67 2,69 (0,571)	68 2,03 (0,304)	68 2,53 (0,446)	67 0,65 (0,064) 0,58 (0,095)	68 0,49 (0,048) 0,42 (0,082)	
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung); SE: <i>standard error</i> (Standardfehler); H'g: Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i></p> <p>a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 96.</p> <p>GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i>, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline der <i>Serumphosphat-Konzentration</i> als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert der <i>Serumphosphat-Konzentration</i> ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry</i> Matrix verwendet.</p>								

Die XLH ist eine seltene, genetisch bedingte, chronische Multisystemerkrankung mit einem komplexen Krankheitsbild, die sich bereits im frühen Kindesalter manifestiert und zu lebenslangen schweren Beeinträchtigungen führt. Das Hauptkennzeichen dieser Erkrankung ist ein chronischer Phosphatverlust, der durch eine vermehrte Expression des Fibroblasten-Wachstumsfaktors FGF23 hervorgerufen wird. Infolgedessen ist die Resorption von Kalzium und Phosphat in das Knochenskelett reduziert und die Knochenmineralisierung beeinträchtigt. Zu den Folgeerkrankungen der XLH zählen Skelettdeformationen, Osteomalazie, Pseudofrakturen, früh einsetzende Arthrose, Enthesopathien und Spinalkanalstenosen. Die damit assoziierten starken chronischen Schmerzen und ausgeprägte Steifheit führen zu einer erheblichen Einschränkung im Alltag der XLH-Patienten.

Das primäre Therapieziel bei der XLH stellt die Normalisierung der Serumphosphat-Konzentration (LLN:  $\geq 2,5$  mg/dl) dar, das jedoch unter der bisher einzigen verfügbaren Option, einer Substitutionstherapie bestehend aus oralem Phosphat und aktivem Vitamin D, nicht erreicht werden konnte. Dies wird anhand der Studienteilnehmer der Studie UX023-CL303 deutlich, bei denen die Serumphosphat-Konzentration (mg/dl) zu Baseline trotz langjähriger Phosphatsubstitution im Durchschnitt unterhalb des unteren Normbereichs von 2,5 mg/dl lag (Burosumab: MW: 2,03 mg/dl, Placebo: MW: 1,92 mg/dl). Durch Burosumab wurde dagegen bereits nach einer 12-wöchigen Behandlung eine durchschnittliche Serumphosphat-Konzentration im unteren Normbereich von 2,69 mg/dl erzielt. In der Placebo-Gruppe betrug die durchschnittliche Serumphosphat-Konzentration lediglich 2,06 mg/dl. Die Wirkung von Burosumab konnte auch im Verlauf bis zu Woche 24 aufrechterhalten werden (Burosumab: MW 2,53 mg/dl, Placebo: MW 2,07 mg/dl). Während in der Burosumab-Gruppe die adjustierte mittlere Änderung (LsMean) der Serumphosphat-Konzentration 0,42 mg/dl betrug, erreichte die Placebo-Gruppe lediglich eine adjustierte mittlere Änderung von 0,03 mg/dl. Der Behandlungsunterschied zwischen den Behandlungsgruppen zu Woche 24 war statistisch signifikant und klinisch relevant (LsMean<sub>w24</sub>: 0,40 mg/dl;  $p < 0,0001$ ,  $H'g = 0,922$  [0,566; 1,278]) (Tabelle 4-30).

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-31: Ergebnisse für *Serumphosphat* – Responderanalyse aus RCT mit dem zu bewertenden Arzneimittel (Woche 24)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	ARR <sup>a,b,c</sup> [95% KI]	p-Wert <sup>a</sup>	RR <sup>b,d</sup> [95% KI]	ARD <sup>b,d</sup> [95% KI]	p-Wert <sup>e</sup>
<i>Anteil der Patienten mit einer Serumphosphat-Konzentration im unteren Normbereich (≥ 2,5 mg/dl)</i>							
Burosumab (68)	63 (92,6)	208,430 [46,646; 931,345]	12,43 [10,44; 12,99]	< 0,0001	12,229 [5,252; 28,479]	0,851 [0,762; 0,940]	< 0,0001
Placebo (66)	5 (7,6)						
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Responder; %: Anteil von N; [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>, RR: <i>Risk Ratio</i> (relatives Risiko) ARR: <i>adjusted Risk Ratio</i> (adjustiertes RR), ARD: <i>Absolute Risk Difference</i> (absolute Risikodifferenz)</p> <p>Responder sind definiert als Studienteilnehmer, die nach 24 Wochen eine Serumphosphat-Konzentration von ≥ 2,5 mg/dl aufweisen.</p> <p>Ein RR, ARR bzw. OR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR, ARR bzw. OR &lt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe, den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) und die Altersgruppen als unabhängige Variablen.</p> <p>b: Berechnung der Effektschätzer (OR, RR, ARR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das adjustierte RR und die dazugehörigen Konfidenzintervalle wurden <i>post-hoc</i> nach der Methode von Zhang und Yu bestimmt (adjustiert für die Behandlungsgruppe, den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) und Altersgruppen (siehe Abschnitt 4.2.5.2) [45]).</p> <p>d: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>e: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>							

Das primäre Therapieziel der XLH-Therapie, eine Serumphosphat-Konzentration im unteren Normbereich (LLN: ≥ 2,5 mg/dl), konnte nach 24 Wochen bei 92,6 % der Studienteilnehmer, die Burosumab erhielten, erreicht werden. Dies entspricht einem Behandlungsunterschied von 85,1 % im Vergleich zur Placebo-Gruppe. Das RR betrug 12,229 (95% KI: [5,252; 28,479]) und erreichte mit p < 0,0001 eine statistische Signifikanz (Tabelle 4-31).

*Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn*

Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 *Brief Pain Inventory (BPI) Schlimmster Schmerz* des patientenberichteten Fragebogens *BPI* und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens *BPI* bestimmt. Die verkürzte Form des Fragebogens enthielt die Fragen 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-32). Das für die Stratifizierung verwendete *Interactive web response system (IWRS)* war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*) verwendet wurde (Anhang 4-E 8b).

Tabelle 4-32: Nummerierung der einzelnen Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	- <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	- <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.
9 A – G	- <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:

a: IWRS: *Interactive web response system*  
b: für die Stratifizierung wurde eine verkürzte Form des BPI-Fragebogens verwendet, der nur die Items der Domäne Schmerzintensität enthielt.

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Dies wurde als Protokollabweichung im *Electronic Data Capture System (EDC)* dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher*

*Schmerz  $\leq 6,0$  bzw. BPI Durchschnittlicher Schmerz  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien BPI Schlimmster Schmerz  $\leq 6,0$  und BPI Schlimmster Schmerz  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (BPI Durchschnittlicher Schmerz  $\leq 6,0$  und BPI Durchschnittlicher Schmerz  $> 6,0$ ) dem schlimmsten Schmerz (BPI Schlimmster Schmerz  $\leq 6,0$  und BPI Schlimmster Schmerz  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist.*

Um einen möglichen Einfluss dieser fehlerhaften Stratifizierung auf den primären Wirksamkeitsendpunkt *Serumphosphat-Prozentsatz erreichter Normwert* zu analysieren, wurde zusätzlich eine Sensitivitätsanalyse durchgeführt, die für den ursprünglich geplanten Stratifizierungsfaktor *BPI Schlimmster Schmerz* adjustierte. Die Ergebnisse waren übereinstimmend mit der Primäranalyse und bestätigen, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse der Studie UX023-CL303 hat. Da keine Abweichungen feststellbar waren, wird diese Analyse nicht im Abschnitt 4.3.1.3.1.1 dargestellt, sondern ausschließlich in Anhang 4-G aufgeführt.

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte, kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

#### 4.3.1.3.1.2 Gehfähigkeit – RCT

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-33: Operationalisierung von Gehfähigkeit

Studie	Operationalisierung
UX023-CL303	<p>Die <i>Gehfähigkeit</i> wurde in der Studie UX023-CL303 anhand des <i>6MWT</i> erhoben. Hierbei muss der Patient auf einer vorab abgemessenen Strecke kontinuierlich für sechs Minuten gehen. Anschließend wird die zurückgelegte Gesamtstrecke in Metern gemessen. Um eine möglichst hohe Validität zu gewährleisten, wird der Test von vorab geschultem klinischen Personal (bevorzugt lizenzierte Physiotherapeuten) auf Basis der <i>American Thoracic Society Guidelines</i> (ATS) durchgeführt [28].</p> <p>Vor Beginn der Studie erhielt jedes Prüfzentrum ein klinisches Handbuch (<i>Clinical Evaluator User Manual</i>), in dem detaillierte Angaben zum Ablauf und zur Auswertung des <i>6MWT</i> enthalten waren [29]. Diese Instruktionen sollten einen einheitlichen Ablauf des <i>6MWT</i> über den gesamten Studienverlauf unabhängig vom Studienzentrum gewährleisten. Die einheitliche Durchführung wird im Folgenden detailliert beschrieben:</p> <p><u>Laufstrecke</u></p> <p>Als Laufstrecke sollte vom Studienzentrum ein schwach frequentierter Flur gewählt werden, der nicht mit einem Teppichboden ausgelegt war und Maße von ca. 30 m Länge und 2,5 m Breite aufwies [29]. Die Laufstrecke wurde vorab mit zwei Pylonen markiert, die in einem Abstand von 20 Metern aufgestellt waren. Die vorgeschriebene Breite des Flurs sollte dabei ein problemloses Wenden um die Pylonen herum gewährleisten, ohne beim Laufen um die Pylonen die Gehgeschwindigkeit der Testpersonen zu beeinflussen. Während des Tests mussten die Studienteilnehmer für sechs Minuten zwischen diesen beiden Pylonen auf- und abgehen, wobei eine Runde aus dem Hinlaufen zur zweiten Pylone und dem Zurücklaufen zum Ausgangspunkt (erste Pylone) bestand (eine Runde = 40 m).</p> <p><u>Probelauf</u></p> <p>Zu Trainingszwecken wurde der <i>6MWT</i> bereits bei der ersten Screeningvisite durchgeführt. Lediglich Studienteilnehmer, die auch an diesem Training teilgenommen hatten, durften während der Studie den <i>6MWT</i> absolvieren.</p> <p><u>Durchführung des 6MWT</u></p> <p>Vor der eigentlichen Durchführung des Tests mussten alle Studienteilnehmer analog zu den ATS-Kriterien sich zunächst zehn Minuten auf einem Stuhl ausruhen [28, 29]. Während dieser Zeit wurde der Gesundheitszustand evaluiert und Anweisungen zum Ablauf des Tests gemacht. Studienteilnehmer, die einen Ruhepuls von über 120 Schlägen pro Minute, einen systolischen Blutdruck von über 180 mm Hg oder einen diastolischen Blutdruck von über 100 mm Hg aufwiesen, wurden von dem Test aus Sicherheitsgründen ausgeschlossen. Orthesen und Gehhilfen waren während des Tests gestattet, Rollstühle durften dagegen nicht verwendet werden.</p> <p>Neben den allgemeinen Testanweisungen demonstrierte das Studienpersonal zusätzlich den Studienteilnehmern die Wendung um die Pylonen. Anschließend wurde der Patient aufgefordert innerhalb von sechs Minuten so schnell wie möglich um die beiden Pylonen zu gehen. Eine Motivation des Studienteilnehmers durch Zurufe während dieser sechs Minuten erfolgte dagegen nicht. Analog zu den ATS-Kriterien wurde lediglich nach jeder Minute dem Studienteilnehmer die verbleibende Zeit mitgeteilt. Sobald die sechs Minuten erreicht waren, wurde der Studienteilnehmer aufgefordert stehen zu bleiben, um die final zurückgelegte Strecke zu messen. Anschließend setzte sich der Studienteilnehmer wieder auf einen Stuhl, sodass nachträgliche Messungen zum Gesundheitszustand des Patienten durchgeführt werden konnten.</p>

Studie	Operationalisierung
	<p><u>Berechnung der zurückgelegten Gesamtdistanz</u></p> <p>Jede zurückgelegte Runde (40 Meter), die sowohl aus dem Hin- als auch dem Zurücklaufen zwischen den beiden Pylonen im Abstand von 20 Metern bestand, wurde durch das lizenzierte Studienpersonal im 6MWT Worksheet anhand einer Strichliste festgehalten. Falls die zuletzt angefangene Runde innerhalb der sechs Minuten nicht vollständig beendet werden konnte, wurde die bis dahin zurückgelegte Distanz gemessen und auf den nächsten vollständigen Meter gerundet. Anschließend wurde die zurückgelegte Gesamtdistanz in Metern berechnet und in den CRF eingetragen.</p> <p>Weitere Details zur Durchführung des Tests und Einweisung der Studienteilnehmer sind dem <i>Clinical Evaluator User Manual</i> zu entnehmen [29].</p> <p>Neben der tatsächlichen durchschnittlichen Gesamtdistanz, die innerhalb von sechs Minuten zurückgelegt wurde, wurde zusätzlich die zurückgelegte Strecke innerhalb der sechs Minuten als Prozentsatz der vorhergesagten Werte auf Grundlage normativer Daten (normiert auf Alter, Geschlecht und Größe) ermittelt. Die vorausgesagte Distanz basiert auf der Publikation von <i>Gibbons et al.</i> und beruht auf einem gesunden Patientenkollektiv, das aus 88 Erwachsenen im Alter zwischen 20 und 80 Jahren besteht (Altersgruppen 20 – 40 Jahre, 41 – 60 Jahre und 61 – 80 Jahren) [30]. Gegenwärtig existieren Untersuchungen zu klinischen Relevanzschwellen (<i>minimally important difference</i>, MID) des 6MWT, die jedoch zu heterogenen Ergebnissen in Abhängigkeit der Patientenpopulation führen. Da spezifisch für XLH bis zum heutigen Tag keine validierten MID's vorliegen, wurde die klinische Relevanz anhand des Hedges'g bestimmt.</p> <p>Die <i>Gehfähigkeit</i> wurde zur Screening-, und Baseline-Visite sowie in Woche 12 und Woche 24 erhoben.</p> <p><u>Analysepopulation:</u> FAS</p> <p><u>Statistisches Modell:</u></p> <ul style="list-style-type: none"> <li>• Auswertungen zur mittleren Änderung der <i>Gehfähigkeit</i> wurden zu Woche 12 und Woche 24 mit einem GEE-Modell bestimmt. Bei dem GEE-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline <i>6MWT</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline im <i>6MWT</i>. Als Kovarianzstruktur wurde eine <i>Compound Symmetry</i> Matrix verwendet.</li> </ul> <p><u>Darstellung im Dossier:</u></p> <ul style="list-style-type: none"> <li>• Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen (Gesamtrate in Meter und Prozentsatz) <ul style="list-style-type: none"> <li>○ Korrigierte Analyse: der fehlerhaft kalkulierte Baseline-Wert eines Studienteilnehmers wurde korrigiert und in die Auswertung einbezogen. Diese Analyse wurde bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 durch den G-BA zum Ableiten des Zusatznutzens von Burosumab herangezogen [1].</li> <li>○ Sensitivitätsanalyse: Analyse enthält nicht den Studienteilnehmer mit dem fehlerhaften Baseline-Wert</li> </ul> </li> </ul> <p>Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier mittels Hedges'g geprüft</p>

Tabelle 4-34: Bewertung des Verzerrungspotenzials für *Gehfähigkeit* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

Der *6MWT* ist ein objektives Messinstrument zur Evaluation der körperlichen Funktion. Die Durchführung erfolgte unter Berücksichtigung der Kriterien der *American Thoracic Society Guidelines* durch lizenzierte Physiotherapeuten, die vorab geschult wurden. Sowohl die Studienteilnehmer als auch das Studienpersonal und der Prüfarzt waren während der 24-wöchigen, doppelblinden, Placebo-kontrollierten Studie der Phase III gegenüber der Behandlung verblindet. Da alle Studienteilnehmer während der 24-wöchigen Phase mindestens einmal die Prüfsubstanz erhielten, umfasst die FAS die gesamte Studienpopulation. Somit fand eine ergebnisunabhängige Berichterstattung unter adäquater Umsetzung des ITT-Prinzips statt.

Zusammenfassend liegt für den Endpunkt *Gehfähigkeit* ein **niedriges Verzerrungspotenzial** vor.

*Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.*

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-35: Ergebnisse für *Gehfähigkeit* aus RCT mit dem zu bewertenden Arzneimittel (korrigierte Analyse)

Intervention	6MWT (6-Minuten-Gehdistanz Test)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MWD (SE) LsMean <sup>a</sup> (SE)	n MWD (SE) LsMean <sup>a</sup> (SE)	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
<b>Korrigierte Analyse<sup>b</sup></b>								
<b>Gesamrate – Messung in Meter</b>								
Placebo (N=66)	65 367,42 (103,408)	64 368,80 (103,619)	64 370,92 (99,567)	64 367,28 (104,220)	64 369,44 (103,385)	64 2,13 (5,026) -0,18 (7,342)	64 2,16 (5,346) -1,65 (7,865)	19,83 [4,19; 35,47] 0,0129 H <sup>g</sup> = 0,424 [0,075; 0,773]
Burosumab (N=68)	67 356,78 (109,464)	64 355,88 (111,604)	64 373,97 (106,108)	65 359,51 (109,629)	65 381,45 (108,464)	64 18,09 (5,214) 13,33 (6,997)	65 21,94 (6,076) 18,19 (7,657)	
<b>Gesamrate – Prozentsatz (%)</b>								
Placebo (N=66)	65 52,25 (14,901)	64 52,44 (14,939)	64 52,71 (14,054)	64 52,16 (15,003)	64 52,42 (14,329)	64 0,27 (0,703) -0,39 (1,002)	64 0,26 (0,737) -0,63 (1,079)	3,19 [0,96; 5,41] 0,0050 H <sup>g</sup> = 0,460 [0,110; 0,809]
Burosumab (N=68)	67 51,43 (15,829)	64 51,31 (16,128)	64 54,02 (15,508)	65 51,87 (15,806)	65 55,19 (15,857)	64 2,72 (0,755) 1,80 (0,983)	65 3,32 (0,889) 2,56 (1,105)	
<p>MWD: Mittelwertsdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung), SE: <i>standard error</i> (Standardfehler); H<sup>g</sup>: Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i></p> <p>a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 96.</p> <p>b: Bei dieser Analyse wurde ein zunächst fehlerhaft kalkulierter Baseline-Wert eines Studienteilnehmers korrigiert und in die Auswertung mit einbezogen.</p> <p>GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i>, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline des <i>6MWT</i> als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI</i></p>								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention	<i>6MWT (6-Minuten-Gehdistanz Test)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>LsMean<sup>a</sup></b>
	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MWD (SE)</b>	<b>MWD (SE)</b>	<b>[95% KI]</b>
	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>p-Wert</b>
								<b>falls p &lt; 0,05:</b>
								<b>Hedges'g [95% KI]</b>
<p><i>Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert des <i>6MWT</i> ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry</i> Matrix verwendet.</p>								

Erwachsene mit XLH weisen aufgrund der Skelettdeformationen in den unteren Extremitäten, der früh einsetzenden Arthrose und Achsenfehlstellungen häufig einen charakteristischen watschelnden Gang sowie Knochen- und Gelenkschmerzen auf. Die chronische Hypophosphatämie führt zudem zu Osteomalazie, Pseudofrakturen, Enthesopathien, Steifheit und verminderter Muskelkraft. Infolgedessen sind die Bewegungsfähigkeit, das Gleichgewicht und die Koordination stark beeinträchtigt. Um den Einfluss von Burosumab auf die Gehfähigkeit zu untersuchen, wurde der *6MWT* durchgeführt als Indikator für die Mobilität und körperliche Funktionsfähigkeit. Hierbei wurde die durchschnittliche Gesamtstrecke in Metern, die innerhalb von sechs Minuten bewältigt werden konnte (absoluter Wert) gemessen. Des Weiteren wurde die innerhalb von sechs Minuten zurückgelegte Strecke als Prozentsatz der vorhergesagten Werte auf Grundlage normativer Daten ermittelt (relativer Wert). Zusätzlich wurde eine Sensitivitätsanalyse durchgeführt, da bei einem Studienteilnehmer der Burosumab-Gruppe aufgrund eines Kalkulationsfehlers durch das Studienzentrum zu Beginn der Studie ein falscher Baseline-Wert in die Datenbank eingetragen wurde, der nach der Auswertung zu Woche 24 korrigiert wurde. In dieser Analyse wurde der Studienteilnehmer mit dem fehlerhaften Baseline-Wert ausgeschlossen, um zu überprüfen, ob die nachträgliche Korrektur des Baseline-Wertes einen Einfluss auf das Ergebnis in der Studie hat.

#### *6MWT absolute Gehdistanz (korrigierte Analyse)*

Die Burosumab-Gruppe wies zu Baseline eine stärkere Beeinträchtigung der Gehfähigkeit als die Placebo-Gruppe auf. Während die Burosumab-Gruppe in sechs Minuten lediglich durchschnittlich 356,78 m zurücklegte, erreichte die Placebo-Gruppe innerhalb dieser Zeit durchschnittlich 367,42 m. Nach 12 Wochen konnte dagegen in der Burosumab-Gruppe bereits eine deutliche Verbesserung der Mobilität gemessen werden: die zurückgelegte Gehstrecke in der Burosumab-Gruppe nahm durchschnittlich um 18,09 m (LsMean: 13,33 m) zu, während die Placebo-Gruppe lediglich durchschnittlich eine Steigerung von 2,13 m (LsMean: -0,18 m) erreichte. Trotz unterschiedlicher Gehdistanz zu Baseline konnte die Burosumab-Gruppe somit bereits nach 12 Wochen eine weitere Strecke zurücklegen als die Placebo-Gruppe. Dieser Trend setzte sich auch in Woche 24 fort. Die adjustierte mittlere Änderung in der Burosumab-Gruppe betrug 18,19 m, in der Placebo-Gruppe verringerte sich die Gehstrecke um 1,65 m. Die Behandlungsdifferenz zwischen den beiden Gruppen von 19,83 m im Vergleich zu Baseline bestätigt den positiven Effekt von Burosumab auf die Gehfähigkeit und war mit  $p = 0,0129$  statistisch signifikant (Tabelle 4-35).

Gegenwärtig existieren Untersuchungen zu klinischen Relevanzschwellen (MID) des *6MWT*, die jedoch zu heterogenen Ergebnissen in Abhängigkeit der Patientenpopulation führen [59, 60]. Da für XLH bis zum heutigen Tag keine validierten MIDs vorliegen, wurde die klinische Relevanz anhand des Hedges'g bestimmt. Der statistisch signifikante Effekt zugunsten von Burosumab erreichte nach 24 Wochen jedoch nicht die klinische Relevanzschwelle ( $H'g = 0,424 [0,075; 0,773]$ ) (Tabelle 4-35)

#### *6MWT relative Gehdistanz (korrigierte Analyse)*

Zusätzlich zu der Gesamtdistanz wurde die innerhalb von sechs Minuten zurückgelegte Strecke als Prozentsatz der vorhergesagten Werte auf Grundlage normativer Daten (normiert auf Alter,

Geschlecht und Größe) ermittelt. Die hierfür verwendeten Referenzwerte stammten von einem gesunden Patientenkollektiv und beziehen sich auf die Altersgruppen 20 – 40 Jahre, 41 – 60 Jahre und 61 – 80 Jahren [30]. Zu Baseline unterschieden sich die relativen Werte zwischen den Behandlungsgruppen nicht (Burosumab: MW: 51,43 %, Placebo: MW: 52,25 %). Nach 12 Wochen zeigte sich dagegen in der Burosumab-Gruppe eine Verbesserung der Gehfähigkeit, die sich bis zu Woche 24 weiter steigerte. Während in der Burosumab-Gruppe der relative Wert um 2,56 Prozentpunkte anstieg, war in der Placebo-Gruppe eine Verschlechterung um 0,63 Prozentpunkte zu verzeichnen. Die Behandlungsdifferenz von 3,19 Prozentpunkten war mit  $p = 0,0050$  statistisch signifikant (Tabelle 4-35).

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-36: Ergebnisse für *Gehfähigkeit* aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse)

Intervention (N)	6-Minuten-Gehdistanz Test (6MWT)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b>Sensitivitätsanalyse<sup>b</sup> (ohne Studienteilnehmer mit fehlerhaftem Baseline-Wert)</b>								
<b>Gesamtrate – Messung in Meter</b>								
Placebo (N=66)	65 367,42 (103,408)	64 368,80 (103,619)	64 370,92 (99,567)	64 367,28 (104,220)	64 369,44 (103,385)	64 2,13 (5,026) -0,42 (7,348)	64 2,16 (5,346) -1,89 (7,866)	21,08 [5,50, 36,65] 0,0080 H <sup>g</sup> = 0,458 [0,107; 0,809]
Burosumab (N=67)	66 355,21 (109,546)	63 354,22 (111,708)	63 372,10 (105,887)	64 357,94 (109,757)	64 381,25 (109,310)	63 17,87 (5,292) 12,69 (6,982)	64 23,31 (6,012) 19,19 (7,640)	
<b>Gesamtrate - Prozentsatz</b>								
Placebo (N=66)	65 52,25 (14,901)	64 52,44 (14,939)	64 52,71 14,054	64 52,16 (15,003)	64 52,42 (14,329)	64 0,27 (0,703) -0,41 (1,002)	64 0,26 (0,737) -0,65 (1,078)	3,36 [1,15; 5,58] 0,0029 H <sup>g</sup> = 0,496 [0,144; 0,848]
Burosumab (N=67)	66 51,17 (15,805)	63 51,03 (16,105)	63 53,72 (15,433)	64 51,61 (15,788)	64 55,13 (15,976)	63 2,68 (0,766) 1,71 (0,980)	64 3,53 (0,879) 2,71 (1,101)	
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung) SE: <i>standard error</i> (Standardfehler); H<sup>g</sup>: Hedges'g (wurde <i>post-hoc</i> berechnet)</p> <p>a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 96.</p> <p>b: Diese Analyse beinhaltet nicht den Studienteilnehmer mit dem fehlerhaft kalkulierten Baseline-Wert.</p> <p>GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i>, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline des <i>6MWT</i> als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI</i></p>								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention (N)	6-Minuten-Gehdistanz Test (6MWT)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>LsMean<sup>a</sup></b>
	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MWD (SE)</b>	<b>MWD (SE)</b>	<b>[95% KI]</b>
	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>p-Wert</b>
								<b>falls p &lt; 0,05: Hedges'g [95% KI]</b>
<p><i>Durchschnittlichen Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert des 6MWT ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry</i> Matrix verwendet.</p>								

*Sensitivitätsanalyse aufgrund einer fehlerhaften Baseline-Wert-Berechnung zu Beginn der Studie*

Bei einem Studienteilnehmer der Burosumab-Gruppe wurde zu Beginn der Studie der Baseline-Wert durch das Studienzentrum falsch kalkuliert. Dieser Fehler ist nach der Primäranalyse zu Woche 24 bemerkt worden und wurde anschließend in der Datenbank korrigiert. Um einen möglichen Einfluss dieser nachträglichen Korrektur auf das Ergebnis des *6MWT* auszuschließen, wurde eine Sensitivitätsanalyse durchgeführt, bei der dieser Studienteilnehmer ausgeschlossen wurde. Durch den Ausschluss des Studienteilnehmers ergab sich kein abweichendes Ergebnis, sodass die korrigierte Analyse zum Ableiten des Zusatznutzens herangezogen wird (Tabelle 4-36).

Zusammenfassend lässt sich feststellen, dass durch Burosumab nicht nur die Serumphosphat-Homöostase innerhalb von 24 Wochen wiederhergestellt wird, sondern auch die Gehfähigkeit signifikant verbessert wird.

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

#### 4.3.1.3.1.3 Steifheit und physische Funktion – RCT

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-37: Operationalisierung von *Steifheit* und *physische Funktion*

Studie	Operationalisierung
UX023-CL303	<p>Der <i>Western Ontario and McMaster Universities Osteoarthritis (WOMAC)</i>-Fragebogen besteht aus insgesamt 24 Items und unterteilt sich in die 3 Domänen <i>WOMAC Steifheit</i> (2 Fragen), <i>WOMAC Physische Funktion</i> (17 Fragen) und <i>WOMAC Schmerzen</i> (5 Fragen). Zum Messen des Ausmaßes von Arthrose auf die Steifheit und physische Funktion wurde die Likert-Skala von 0 bis 4 verwendet (keine, milde, moderate, schwere und extreme Auswirkung). Dabei werden für jede Domäne die Werte der Antwortmöglichkeiten addiert, sodass sich verschiedene Spannweiten für die einzelnen Domänen ergeben (<i>Steifheit</i> 0 – 8 Punkte, <i>Physische Funktion</i> 0 – 68 Punkte, Schmerz 0 – 20 Punkte). Ein höherer Score indiziert dabei eine höhere Ausprägung an Steifheit sowie funktioneller Beeinträchtigung bzw. Schmerz. Die Scores werden anschließend auf einer metrischen Skala von 0 – 100 (<i>normalised units</i>, normierte Einheiten, nu) normiert, wobei 0 den besten Gesundheitszustand reflektiert und 100 den schlechtesten [31]. Das gleiche Vorgehen wurde auch bei der Bildung des Gesamtwerts angewendet.</p> <p>Die im Studienprotokoll präspezifizierten Responderanalysen klassifizierten Patienten mit einer Abnahme des Scores um mindestens 9,3 nu (<i>normalised units</i>, nu, <i>WOMAC Physische Funktion</i>) und bei der Domäne <i>WOMAC Steifheit</i> eine Abnahme des Scores um mindestens 10 nu als Responder [34, 35]. Für das vorliegende Dossier werden jedoch analog zu den Anforderungen im Methodenpapier des IQWiG jeweils eine Responderchwelle mit einer Skalenspannweite von mindestens 15 % (entspricht <math>\geq 15</math> nu) herangezogen [8].</p> <p>Die Domänen <i>WOMAC Steifheit</i>, <i>WOMAC Physische Funktion</i> und <i>WOMAC Schmerz</i> wurden zur Baseline-Visite, zu Woche 12 und Woche 24 erhoben.</p> <p>Die Domäne <i>WOMAC Schmerz</i> wurde in der Studie UX023-CL303 zwar erhoben, aber nicht für die Auswertung berücksichtigt, da der Schmerz bereits durch den Endpunkt <i>Brief Pain Inventory</i> (BPI) beurteilt wurde (eine detaillierte Begründung ist in Abschnitt 4.2.5.2 aufgeführt).</p> <p><u>Analysepopulation:</u> FAS</p> <p><u>Statistisches Modell:</u></p> <ul style="list-style-type: none"> <li>• Auswertungen zur mittleren Änderung der <i>Steifheit</i>, <i>physischen Funktion</i> wurden zu Woche 12 und Woche 24 mit einem GEE-Modell bestimmt. Bei dem GEE-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline in <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion</i>. Als Kovarianzstruktur wurde eine <i>Compound Symmetry</i> Matrix verwendet. Analog hierzu wurde ebenfalls der <i>WOMAC Gesamtscore</i> und <i>WOMAC Schmerz</i> berechnet.</li> <li>• Zusätzlich wurde <i>post-hoc</i> eine Auswertung der mittleren Änderung der <i>Steifheit</i> und <i>Physischen Funktion</i> zu Woche 12 und Woche 24 mit einem MMRM-Modell durchgeführt. Beim MMRM-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline in <i>WOMAC Steifheit</i> bzw. <i>WOMAC</i></li> </ul>

*Physische Funktion.* Als Kovarianzstruktur wurde eine unstrukturierte Kovarianz (*WOMAC Steifheit*) bzw. bei nicht-erfolgter Konvergenz des Modells eine heterogene Toeplitz Kovarianzstruktur (*WOMAC Physische Funktion*) verwendet.

- Responderanalyse Anteil an Studienteilnehmern mit einer Verbesserung der *Steifheit* bzw. der *Physischen Funktion* um  $\geq 15$  nu: Für das Dossier wurde *post-hoc* das OR inklusive 95% KI und p-Wert anhand eines logistischen Regressionsmodells analysiert (Woche 24). Die Behandlungsgruppe und der tatsächliche Stratifizierungsfaktor (*BPI durchschnittlicher Schmerz*) wurden als unabhängige Variablen in das Modell aufgenommen. Das adjustierte RR (Adjustierung analog zum OR) und die dazugehörigen Konfidenzintervalle wurden *post-hoc* nach der Methode von Zhang und Yu bestimmt (adjustiert für die Behandlungsgruppe und den tatsächlichen Stratifizierungsfaktor (*BPI Durchschnittlicher Schmerz*)) (siehe Abschnitt 4.2.5.2) [45]. Aufgrund der geringen Anzahl an Studienteilnehmern aus Japan ( $n = 11$ ) und Südkorea ( $n = 7$ ) und folglich einer daraus resultierenden geringen Anzahl an Respondern, konnte der Faktor Region nicht, wie ursprünglich im SAP vorgesehen, als unabhängige Variable im logistischen Regressionsmodell berücksichtigt werden. Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation und der p-Wert mittels Fishers exaktem Test berechnet.

#### Darstellung im Dossier:

- *WOMAC Steifheit*
  - Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen
  - Responderanalyse: Anteil an XLH-Patienten mit einer Abnahme von  $\geq 15$  nu zu Woche 24
  - Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn: Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen adjustiert nach dem eigentlichen Stratifizierungsfaktor *BPI Schlimmster Schmerz*
  - Sensitivitätsanalysen aufgrund fehlender Post-Baseline-Werte: mBOCF (*modified baseline observation carried forward*), BOCF (*baseline observation carried forward*), LOCF (*last observation carried forward*)
  - Sensitivitätsanalyse: Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen ausgewertet mittels MMRM
  - Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier anhand von Hedges'g geprüft
- *WOMAC Physische Funktion*
  - Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen
  - Responderanalyse: Anteil an XLH-Patienten mit einer Abnahme von  $\geq 15$  nu zu Woche 24
  - Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn: Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen adjustiert nach dem eigentlichen Stratifizierungsfaktor *BPI Schlimmster Schmerz*
  - Sensitivitätsanalysen aufgrund fehlender Post-Baseline-Werte: mBOCF (*modified baseline observation carried forward*), BOCF (*baseline observation carried forward*), LOCF (*last observation carried forward*)
  - Sensitivitätsanalyse: Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen ausgewertet

Studie	Operationalisierung
	mittels MMRM <ul style="list-style-type: none"> <li>○ Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier anhand von Hedges' g geprüft</li> <li>• <i>WOMAC Schmerz und WOMAC Gesamtscore</i> <ul style="list-style-type: none"> <li>○ Der inhaltlichen Vollständigkeit halber werden die Analysen im Anhang 4-G aufgeführt.</li> </ul> </li> </ul>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-38: Bewertung des Verzerrungspotenzials für *Steifheit und Physische Funktion* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Der *Western Ontario and McMaster Universities Osteoarthritis (WOMAC)*-Fragebogen ist ein selbsterklärender patientenberichteter Fragebogen, dessen Inhaltsvalidität anhand der Leitlinien der FDA überprüft wurde und gewährleistet ist. Sowohl die Studienteilnehmer als auch das Studienpersonal und der Prüfarzt waren während der 24-wöchigen, doppelblinden, Placebo-kontrollierten Phase der Studie UX023-CL303 gegenüber der Behandlung verblindet. Da alle Studienteilnehmer während der 24-wöchigen Phase mindestens einmal die Prüfsubstanz erhielten, umfasst die FAS die gesamte Studienpopulation. Somit erfolgte eine ergebnisunabhängige Berichterstattung unter adäquater Umsetzung des ITT-Prinzips.

Aufgrund der zu Beginn fehlerhaften durchgeführten Stratifizierung (Abschnitt 4.2.5.4, Anhang 4-F) wurden, um die Robustheit der Ergebnisse zu überprüfen, Sensitivitätsanalysen für die sekundären Wirksamkeitsendpunkte **WOMAC Steifheit** und **WOMAC Physische Funktion** durchgeführt. Diese Analysen waren jedoch übereinstimmend mit den Primäranalysen und bestätigen daher, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse der Studie UX023-CL303 hat.

Der Rücklauf des Fragebogens in der Studie belief sich zu allen Messzeitpunkten auf über 70 %.

Zusammenfassend liegt für den Endpunkt *Steifheit und Physische Funktion* in der randomisierten, doppelblinden, Placebo-kontrollierten Studie UX023-CL303 ein **niedriges Verzerrungspotenzial** vor.

*Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.*

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-39: Ergebnisse der *Steifheit* und *Physischen Funktion* aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	WOMAC ( <i>Western Ontario and McMaster Universities Osteoarthritis</i> )							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b>WOMAC Steifheit</b>								
Placebo (N=66)	66 61,36 (20,770)	66 61,36 (20,770)	66 59,28 (19,772)	65 61,15 (20,861)	65 60,38 (21,827)	66 -2,08 (2,186) -0,86 (2,784)	65 -0,77 (2,698) 0,46 (3,319)	-8,31 [-14,68, -1,94] 0,0106 H <sup>g</sup> = -0,420 [-0,765; -0,075]
Burosumab (N=68)	68 64,71 (20,253)	68 64,71 (20,253)	68 54,23 (25,700)	67 64,37 (20,209)	67 53,73 (20,759)	68 -10,48 (2,792) -7,72 (3,533)	67 -10,63 (2,987) -7,85 (3,034)	
<b>WOMAC Physische Funktion</b>								
Placebo (N=66)	66 43,89 (19,938)	66 43,89 (19,938)	66 40,49 (21,386)	65 43,62 (19,967)	65 42,65 (22,760)	66 -3,41 (1,729) -0,67 (2,633)	65 -0,97 (1,826) 1,79 (2,722)	-4,90 [-9,76, -0,05] 0,0478 <sup>b</sup> H <sup>g</sup> = -0,389 [-0,735; -0,043]
Burosumab (N=68)	68 50,79 (19,660)	68 50,79 (19,660)	68 43,23 (20,311)	66 50,33 (19,336)	66 43,43 (19,507)	68 -7,56 (1,377) -3,75 (2,075)	66 -6,90 (1,886) -3,11 (2,553)	
MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung); SE: <i>standard error</i> (Standardfehler); H <sup>g</sup> : Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i> a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 24. b: Nach einer Hochberg-Adjustierung für multiples Testen, die laut CSR präspezifiziert war, wurde aufgrund des herabgesetzten Signifikanzniveaus (p<0,025) keine statistische Signifikanz erreicht. Laut dem Methodenpapier des IQWiG muss bei der Nutzenbewertung jedoch nicht für multiples Testen adjustiert werden, sodass dies für die vorliegende Nutzenbewertung nicht von Bedeutung ist [8].								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention (N)	WOMAC ( <i>Western Ontario and McMaster Universities Osteoarthritis</i> )							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>LsMean<sup>a</sup></b>
	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MWD (SE)</b>	<b>MWD (SE)</b>	<b>[95% KI]</b>
	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>p-Wert</b>
								<b>falls p &lt; 0,05: Hedges'g [95% KI]</b>

GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts *BPI Durchschnittlichen Schmerz*, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline des Endpunkts *WOMAC Steifheit* bzw. *WOMAC Physische Funktion* als abhängige Variable und Region, Visite, Behandlungsarm, *BPI Durchschnittlicher Schmerz* (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert des Endpunkts *WOMAC Steifheit* bzw. *WOMAC Physische Funktion* ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine *Compound Symmetry* Matrix verwendet.

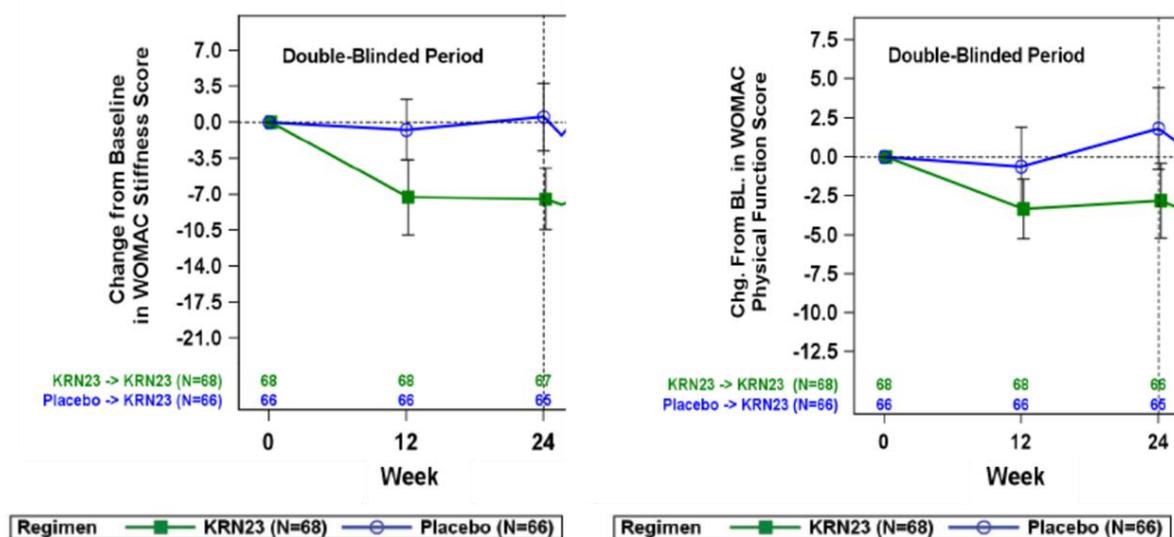


Abbildung 3: Mittelwerte für die Endpunkte *WOMAC Steifheit* und *WOMAC Physische Funktion* über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303

Der *WOMAC*-Fragebogen zählt zu den bei Hüft- und Kniearthrosen am häufigsten verwendeten krankheitsspezifischen Messinstrumenten und wird durch das Komitee von *Outcome Measures in Rheumatology Arthritis Clinical Trials* (OMERACT) bei Arthrose empfohlen [31, 32]. Ziel dieses patientenberichteten Fragebogens ist es, spezifisch die durch Arthrose in den unteren Extremitäten hervorgerufenen Dysfunktionen und Schmerzen zu erfassen. Der Fragebogen umfasst die drei Domänen *Steifheit*, *Physische Funktion* und *Schmerz*. Die Werte werden anhand einer metrischen Skala von 0–100 normiert, wobei ein höherer Score einen schlechteren Gesundheitszustand reflektiert, null dagegen den besten Gesundheitszustand. Da die Domäne *WOMAC Schmerz* sowie der *WOMAC Gesamtscore post-hoc* analysiert wurden, werden die Ergebnisse nicht im Abschnitt 4.3.1.3.1.3 dargestellt, sondern ausschließlich in Anhang 4-G aufgeführt.

#### Domäne *WOMAC Steifheit*

Zu den Ursachen der stark ausgeprägten Steifheit von XLH-Patienten zählen die durch den chronischen Phosphatverlust bedingten Skelettdeformationen, (Pseudo-)Frakturen, Enthesopathien und die früh einsetzende Arthrose. Chronisch niedrige Serumphosphat-Konzentrationen führen zudem zu einer unzureichenden Versorgung der Muskulatur mit ATP, wodurch die Entspannung der Muskulatur beeinträchtigt wird und die Versteifung des Muskel-Sehnen-Apparates gefördert wird [61, 62]. Infolgedessen sind die Muskelfunktion und die körperliche Beweglichkeit stark eingeschränkt.

Der Mittelwert der Domäne *WOMAC Steifheit* betrug in der Burosumab-Gruppe zu Baseline 64,71 und in der Placebo-Gruppe 61,36. Dies entspricht einem etwas stärkeren Ausmaß an Steifheit, als gewöhnlich bei Patienten mit Arthrose zu beobachten ist, und spiegelt die ausgeprägte rheumatische Manifestation von XLH in der Studienpopulation wider [63]. Schon bereits nach 12 Wochen zeigte sich unter der Burosumab-Therapie eine Verbesserung der Steifheit. Während die adjustierte mittlere Änderung (LsMean) in der Burosumab-Gruppe -7,72

betrug, konnte in der Placebo-Gruppe lediglich eine adjustierte mittlere Änderung von -0,86 erzielt werden. Zu Woche 24 konnte die Verringerung der Steifheit in der Burosumab-Gruppe weiterhin aufrechterhalten werden (LsMean: -7,85), während sich in der Placebo-Gruppe das Ausmaß der Steifheit leicht verschlechterte (LsMean: 0,46). Dies wird auch anhand der Behandlungsdifferenz zu Woche 24 ersichtlich, die mit -8,31 statistisch signifikant ( $p = 0,0106$ ) jedoch nicht klinisch relevant ( $H'g = -0,420 [-0,765; -0,075]$ ) war (Tabelle 4-39, Abbildung 3).

*Domäne WOMAC Physische Funktion:*

Obwohl die Burosumab-Gruppe (MW: 50,79) zu Baseline eine stärkere Beeinträchtigung der physischen Funktion aufwies (Placebo: MW: 43,89), konnte nach 12 Wochen nur noch ein geringer Unterschied zwischen den beiden Gruppen festgestellt werden (Burosumab: MW: 43,23; Placebo: MW: 40,49). Während die adjustierte mittlere Änderung in der Burosumab-Gruppe -3,75 betrug, wurde in der Placebo-Gruppe lediglich eine adjustierte mittlere Änderung von -0,67 gemessen. Zu Woche 24 nahm die Beeinträchtigung der physischen Funktion in der Burosumab-Gruppe (LsMean: -3,11) weiterhin ab, in der Placebo-Gruppe war dagegen eine Verschlechterung der physischen Funktion (LsMean: 1,79) zu verzeichnen. Die Behandlungsdifferenz zu Woche 24 von -4,90 unterstreicht die Überlegenheit von Burosumab und war mit  $p = 0,0478$  statistisch signifikant jedoch nicht klinisch relevant ( $H'g = -0,389 [-0,735; -0,043]$ ) (Tabelle 4-39, Abbildung 3).

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-40: Ergebnisse für *Steifheit und Physische Funktion* – Responderanalyse aus RCT mit dem zu bewertenden Arzneimittel (Woche 24)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	ARR [95% KI]	p-Wert <sup>a</sup>	RR <sup>b,d</sup> [95% KI]	ARD <sup>b,d</sup> [95% KI]	p-Wert <sup>e</sup>
<b>WOMAC Steifheit ≥ 15,0 nu</b>							
Burosumab (68)	21 (30,9)	2,519 [1,070; 5,931]	2,09 [1,08; 3,46]	0,0347	2,038 [1,040; 3,994]	0,157 [0,018; 0,297]	0,0403
Placebo (66)	10 (15,2)						
<b>WOMAC Physische Funktion ≥ 15,0 nu</b>							
Burosumab (68)	14 (20,6)	1,259 [0,514; 3,084]	n.b.	0,6127	1,235 [0,605; 2,521]	0,039 [-0,092; 0,171]	0,6591
Placebo (66)	11 (16,7)						
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Responder; %: Anteil von N; [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>, RR: <i>Risk Ratio</i> (relatives Risiko); ARR: <i>adjusted Risk Ratio</i> (adjustiertes relatives Risiko); ARD: <i>Absolute Risk Difference</i> (absolute Risikodifferenz); n.b.: nicht bestimmt</p> <p>Responder sind definiert als Studienteilnehmer, die nach 24 Wochen eine Abnahme des <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion Scores</i> um ≥ 15 % aufweisen.</p> <p>Ein RR, ARR bzw. OR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR, ARR bzw. OR &lt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe und den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) als unabhängige Variablen.</p> <p>b: Berechnung der Effektschätzer (OR, RR, ARR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das adjustierte RR und die dazugehörigen Konfidenzintervalle wurden <i>post-hoc</i> nach der Methode von Zhang und Yu bestimmt (adjustiert für die Behandlungsgruppe und den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) (siehe Abschnitt 4.2.5.2) [45].</p> <p>d: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>e: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>							

*Responderanalysen des WOMAC*

Eine klinisch bedeutsame Verbesserung der Steifheit ( $\geq 15$  nu) wurde zu Woche 24 bei 30,9 % der Studienteilnehmer der Burosumab-Gruppe erzielt. In der Placebo-Gruppe betrug der Anteil dagegen nur 15,2 %. Der Unterschied war statistisch signifikant (RR [95% KI]: 2,038 [1,040; 3,994],  $p = 0,0403$ ). Bei der Domäne *WOMAC Physische Funktion* erlangten 20,6 % der Studienteilnehmer der Burosumab-Gruppe und 16,7 % der Placebo-Gruppe eine klinisch bedeutsame Verbesserung von mindestens 15 nu (RR [95% KI]: 1,235 [0,605; 2,521],  $p = 0,6591$ ) (Tabelle 4-40)

*Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn*

Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 *Brief Pain Inventory (BPI) Schlimmster Schmerz* des patientenberichteten Fragebogens *BPI* und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens *BPI* bestimmt. Die verkürzte Form des Fragebogens enthielt die Fragen 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-41). Das für die Stratifizierung verwendete *Interactive web response system (IWRS)* war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*) verwendet wurde (Anhang 4-E 8b).

Tabelle 4-41: Nummerierung der einzelnen Items des *BPI-Fragebogens* für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	- <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	- <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
9 A – G	_ <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:
<p>a: IWRS: <i>Interactive web response system</i></p> <p>b: für die Stratifizierung wurde eine verkürzte Form des BPI-Fragebogens verwendet, der nur die Items der Domäne Schmerzintensität enthielt.</p>		

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist.

Um einen möglichen Einfluss der fehlerhaften Stratifizierung auf die sekundären Wirksamkeitsendpunkte *WOMAC Steifheit* und *WOMAC Physische Funktion* zu analysieren, wurden Sensitivitätsanalysen durchgeführt, die für den ursprünglich geplanten Stratifizierungsfaktor *BPI Schlimmster Schmerz* adjustieren. Diese Analysen waren jedoch übereinstimmend mit den Primäranalysen und bestätigen daher, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse der Studie UX023-CL303 hat. Da keine Abweichungen feststellbar sind, wird diese Analyse nicht im Abschnitt 4.3.1.3.1.3 dargestellt.

#### *Sensitivitätsanalysen aufgrund fehlender Post-Baseline-Werte*

Um den Einfluss von fehlenden Post-Baseline Werten auf die Ergebnisse zu überprüfen, erfolgten für die beiden sekundären Wirksamkeitsendpunkte *WOMAC Steifheit* und *WOMAC Physische Funktion* zusätzliche Sensitivitätsanalysen (LOCF, mBOCF, BOCF). Diese Analysen waren jedoch übereinstimmend mit den Primäranalysen, wodurch bestätigt wird, dass eine Imputation von Werten keinen Einfluss auf die Aussagekraft der Ergebnisse hat. Daher werden diese Auswertungen nicht im Abschnitt 4.3.1.3.1.3 dargestellt, sondern ausschließlich in Anhang 4-G aufgeführt.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-42: Ergebnisse der *Steifheit* und *Physischen Funktion* aus RCT mit dem zu bewertenden Arzneimittel (MMRM-Analyse)

Intervention (N)	WOMAC ( <i>Western Ontario and McMaster Universities Osteoarthritis</i> ) - MMRM							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b>WOMAC Steifheit</b>								
Placebo (N=66)	66 61,36 (20,770)	66 61,36 (20,770)	66 59,28 (19,772)	65 61,15 (20,861)	65 60,38 (21,827)	66 -2,08 (2,186) -0,87 (3,510)	65 -0,77 (2,698) 0,44 (3,555)	-8,32 [-14,97, -1,67] 0,0146
Burosumab (N=68)	68 64,71 (20,253)	68 64,71 (20,253)	68 54,23 (25,700)	67 64,37 (20,209)	67 53,73 (20,759)	68 -10,48 (2,792) -7,75 (3,365)	67 -10,63 (2,987) -7,88 (3,408)	H <sup>g</sup> = -0,429 [-0,772; -0,086]
<b>WOMAC Physische Funktion</b>								
Placebo (N=66)	66 43,89 (19,938)	66 43,89 (19,938)	66 40,49 (21,386)	65 43,62 (19,967)	65 42,65 (22,760)	66 -3,41 (1,729) -1,10 (2,442)	65 -0,97 (1,826) 1,37 (2,581)	-4,97 [-9,92; -0,02] 0,0491
Burosumab (N=68)	68 50,79 (19,660)	68 50,79 (19,660)	68 43,23 (20,311)	66 50,33 (19,336)	66 43,43 (19,507)	68 -7,56 (1,377) -4,25 (2,346)	66 -6,90 (1,886) -3,60 (2,492)	H <sup>g</sup> = -0,345 [-0,688; -0,002]
MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung); SE: <i>standard error</i> (Standardfehler); H <sup>g</sup> : Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i> a: Berechnung mit MMRM-Modell zu Woche 24  MMRM: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i> , berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem MMRM-Modell mit der Änderung von Baseline des Endpunkts <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion</i> als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. sowie der								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention (N)	WOMAC ( <i>Western Ontario and McMaster Universities Osteoarthritis</i> ) - MMRM							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup>
	MW	MW	MW	MW	MW	MWD (SE)	MWD (SE)	[95% KI]
	(SD)	(SD)	(SD)	(SD)	(SD)	LsMean <sup>a</sup> (SE)	LsMean <sup>a</sup> (SE)	p-Wert
								<b>falls p &lt; 0,05: Hedges'g [95% KI]</b>
Baseline <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline in <i>WOMAC Steifheit</i> bzw. <i>WOMAC Physische Funktion</i> . Als Kovarianzstruktur wurde eine unstrukturierte Kovarianz ( <i>WOMAC Steifheit</i> ) bzw. eine heterogene Toeplitz Kovarianzstruktur ( <i>WOMAC Physische Funktion</i> ) verwendet.								

*Sensitivitätsanalyse MMRM*

Laut der aktuellen Dossiervorlage wird für die Nutzenbewertung empfohlen, für die Auswertung von mit Skalen erhobenen patientenberichteten Endpunkten das MMRM zu verwenden, um die über den Studienverlauf ermittelten Informationen vollständig zu berücksichtigen. Im SAP der Studie UX023-CL303 war jedoch für die Auswertung von patientenberichteten Endpunkten die Verwendung des GEE-Modells präspezifiziert. Ähnlich wie das MMRM erlaubt auch die Verwendung des GEE-Modells die Berücksichtigung patientenspezifischer Heterogenität und stellt daher insbesondere bei Messwiederholungen ein adäquates Auswertungsmodell dar.

Um zu überprüfen, ob die Wahl des statistischen Modells gegebenenfalls einen Einfluss auf die Ergebnisse der patientenberichteten Endpunkte hat, wurde für die Domäne *WOMAC Steifheit* und *WOMAC Physische Funktion* zusätzlich zu der Auswertung mittels GEE eine Analyse mittel MMRM durchgeführt. Diese Analyse war übereinstimmend mit der Primäranalyse (Tabelle 4-42). Aufgrund der übereinstimmenden Ergebnisse, unabhängig von der Wahl des statistischen Modells (MMRM oder GEE-Modell), wird daher im Folgenden für alle patientenberichteten Endpunkte ausschließlich das GEE-Modell verwendet, da zusätzliche Analysen mittels MMRM zu keinen neuen Erkenntnissen führen würden.

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte, kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-

Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

#### 4.3.1.3.1.4 Schmerzzerhebung – RCT

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-43: Operationalisierung von Schmerzzerhebung

Studie	Operationalisierung
UX023-CL303	<p>Die Schmerzzerhebung erfolgte anhand der Kurzform des <i>Brief Pain Inventory (BPI)</i>, einem patientenberichteten Fragebogen, der den Schmerz innerhalb der letzten 24 Stunden abfragt. In dem Fragebogen wird anhand von vier Items die Schmerzintensität mit einer 11-Punkte-Skala (0 = kein Schmerz bis 10 = Schlimmster Schmerz, den man sich vorstellen kann) ermittelt (Frage 3 – 6: <i>BPI Schlimmster Schmerz</i>, <i>BPI Durchschnittlicher Schmerz</i>, <i>BPI Geringster Schmerz</i> und <i>BPI Derzeitiger Schmerz</i>). Zusätzlich wird die Schmerzinterferenz anhand von 7 Items adressiert (9A – 9G). Die Auswertung erfolgt ebenfalls anhand einer 11-Punkte-Skala und reicht von 0 (keine Interferenz) bis 10 (stärkste Interferenz). Die Schmerzinterferenz bezieht sich hierbei auf die Beeinträchtigung bei der Ausführung von Alltagsaktivitäten (Gehfähigkeit, Schlaf, Stimmung, Lebensqualität, generelle Aktivitäten, Arbeit und Beziehungen zu anderen). In der Studie UX023-CL303 wurde die Frage 3 <i>BPI Schlimmster Schmerz</i> als sekundärer Endpunkt separat erhoben, da diese den ursprünglichen Stratifizierungsfaktor darstellt. Die anderen Items wurden dagegen zusätzlich deskriptiv ausgewertet.</p> <p>Responder wurden definiert als Patienten, die bei dem Endpunkt <i>BPI Schlimmster Schmerz</i> eine Abnahme um mindestens 15 % aufwiesen. Eine Abnahme um 30 % stellt dagegen eine für den Patienten bedeutsame Linderung des chronischen Schmerzes dar [64].</p> <p>Die Items der Domäne <i>BPI Schmerzintensität</i> wurden in den 7 Tagen vor der jeweiligen Studienvisite (Baseline, Woche 12 und 24) täglich durch ein Patiententagebuch erfasst und ebenfalls am Tag der Studienvisite erhoben. Anschließend wurde der Mittelwert der 8-tägigen Messung gebildet. Die Items der Domäne <i>BPI Schmerzinterferenz</i> wurden dagegen während der Studienvisite (Baseline, Woche 12 und 24) erhoben (siehe Abschnitt 4.2.5.2).</p> <p><u>Analysepopulation:</u> FAS</p> <p><u>Statistisches Modell:</u></p> <ul style="list-style-type: none"> <li>• Auswertungen zur mittleren Änderung der Endpunkte <i>BPI Schlimmster Schmerz</i>, <i>BPI Schmerzintensität</i> und <i>BPI Schmerzinterferenz</i> wurden zu Woche 12 und Woche 24 mit einem GEE-Modell bestimmt. Bei dem GEE-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline <i>BPI Schlimmster Schmerz</i> bzw. <i>BPI Schmerzintensität</i> oder <i>BPI Schmerzinterferenz</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline in <i>BPI Schlimmster Schmerz</i> bzw. <i>BPI Schmerzintensität</i> oder <i>BPI Schmerzinterferenz</i>. Als Kovarianzstruktur wurde eine <i>Compound Symmetry</i> Matrix verwendet.</li> <li>• Responderanalyse Anteil an Studienteilnehmern mit einer Abnahme des <i>BPI Schlimmsten Schmerzes</i> um 15 % bzw. 30 %: Prospektiv geplant und im Studienbericht der Studie UX023-CL303 dargestellt ist die Auswertung des <i>Odds Ratio (OR)</i> mittels Cochran-Mantel-Haenszel-(CMH-)Test. Für das Dossier wurde <i>post-hoc</i> das OR inklusive 95% KI und p-Wert anhand eines logistischen Regressionsmodells analysiert (Woche 24). Die Behandlungsgruppe, der tatsächliche Stratifizierungsfaktor (<i>BPI durchschnittlicher Schmerz</i>) und die Altersgruppen wurden als unabhängige Variablen in das Modell aufgenommen. Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels</li> </ul>

Studie	Operationalisierung
	<p>Normalverteilungsapproximation und der p-Wert mittels Fishers exaktem Test berechnet.</p> <p><u>Darstellung im Dossier:</u></p> <ul style="list-style-type: none"> <li>• <u>BPI Schlimmster Schmerz</u> <ul style="list-style-type: none"> <li>○ Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen</li> <li>○ Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier anhand von Hedges`g geprüft</li> <li>○ Responderanalyse: Anteil an XLH-Patienten mit einer Abnahme des <i>BPI Schlimmsten Schmerz</i> von <math>\geq 15\%</math> und <math>\geq 30\%</math> zu Woche 24</li> <li>○ Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn: Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen adjustiert nach dem eigentlichen Stratifizierungsfaktor <i>BPI Durchschnittlicher Schmerz</i></li> <li>○ Sensitivitätsanalysen aufgrund fehlender Post-Baseline-Werte: mBOCF (<i>modified baseline observation carried forward</i>), BOCF (<i>baseline observation carried forward</i>), LOCF (<i>last observation carried forward</i>)</li> </ul> </li> <li>• <u>BPI Schmerzintensität</u> <ul style="list-style-type: none"> <li>○ Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen</li> <li>○ Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier anhand von Hedges`g geprüft</li> </ul> </li> <li>• <u>BPI Schmerzinterferenz</u> <ul style="list-style-type: none"> <li>○ Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier anhand von Hedges`g geprüft</li> </ul> </li> </ul>

*Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.*

*Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.*

Tabelle 4-44: Bewertung des Verzerrungspotenzials für *Schmerzerhebung* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Der *Brief Pain Inventory (BPI)*-Fragebogen ist ein patientenberichteter Fragenbogen, der kein Training der Studienteilnehmer erforderte. Sowohl die Studienteilnehmer als auch das Studienpersonal und der Prüfarzt waren während der 24-wöchigen, doppelblinden, Placebo-kontrollierten Phase der Studie gegenüber der Behandlung verblindet. Da alle Studienteilnehmer während der 24-wöchigen Phase mindestens einmal die Prüfsubstanz erhielten, umfasst die FAS die gesamte Studienpopulation. Somit fand eine ergebnisunabhängige Berichterstattung unter adäquater Umsetzung des ITT-Prinzips statt.

Aufgrund der zu Beginn fehlerhaften durchgeführten Stratifizierung (Abschnitt 4.2.5.4, Anhang 4-F) wurden, um die Robustheit der Ergebnisse zu überprüfen, Sensitivitätsanalysen für den sekundären Wirksamkeitsendpunkt **BPI Schlimmster Schmerz** durchgeführt. Diese Analysen waren jedoch übereinstimmend mit den Primäranalysen und bestätigten daher, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse der Studie UX023-CL303 hat.

Der Rücklauf des Fragebogens in der Studie belief sich zu allen Messzeitpunkten auf über 70 %.

Zusammenfassend liegt für den Endpunkt **Schmerzerhebung** (*BPI Schlimmster Schmerz, BPI Schmerzinterferenz und BPI Schmerzintensität*) in der randomisierten doppelblinden Placebo-kontrollierten Studie UX023-CL303 ein **niedriges Verzerrungspotenzial** vor.

Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-45: Ergebnisse für *Schmerzerhebung (BPI Schlimmster Schmerz)* aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	<i>BPI (Brief Pain Inventory)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b><i>BPI Schlimmster Schmerz</i></b>								
Placebo (N=66)	66 6,54 (1,433)	66 6,54 (1,433)	66 6,03 (1,955)	65 6,51 (1,417)	65 6,09 (2,013)	66 -0,52 (0,192) -0,38 (0,207)	65 -0,42 (0,218) -0,32 (0,222)	-0,46 [-1,00; 0,08] 0,0919
Burosumab (N=68)	68 6,81 (1,308)	68 6,81 (1,308)	68 5,99 (1,762)	67 6,80 (1,316)	67 5,82 (1,916)	68 -0,82 (0,177) -0,63 (0,194)	67 -0,98 (0,191) -0,79 (0,211)	-
MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung); SE: <i>standard error</i> (Standardfehler); H'g: Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i> a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 24. GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i> , berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit dem Änderung von Baseline des Endpunkts <i>BPI Schlimmster Schmerz</i> als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert des Endpunkts <i>BPI Schlimmster Schmerz</i> ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry</i> Matrix verwendet.								

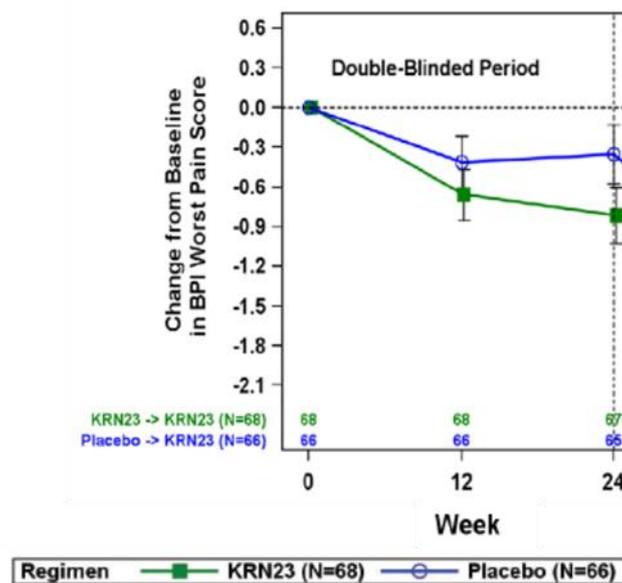


Abbildung 4: Mittelwerte für den Endpunkt *BPI Schlimmster Schmerz* über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303

Aufgrund der Skelettdeformationen, der früh einsetzenden Arthrose und damit verbundener Steifheit, den Enthesopathien, der Osteomalazie und den (Pseudo-)Frakturen leiden XLH-Patienten an stark ausgeprägten chronischen Knochen- und Gelenkschmerzen, die zu einer erheblichen Beeinträchtigung des Alltags führen. Um den Einfluss der Burosumab-Therapie auf die XLH-assoziierten Schmerzen zu analysieren, wurde daher der *Brief Pain Inventory* (BPI) eingesetzt, ein patientenberichteter Fragebogen, der sowohl den *BPI Schlimmsten Schmerz*, die *BPI Schmerzintensität* als auch die *BPI Schmerzinterferenz* anhand einer 11-Punkte Skala misst. Ein höherer Wert zeigt hierbei eine größere Beeinträchtigung bzw. stärkere Schmerzen an.

#### *BPI Schlimmster Schmerz*

Die starke Ausprägung von XLH-bedingten Schmerzen wird anhand der Baseline-Werte des *BPI Schlimmsten Schmerzes* zu Studienbeginn ersichtlich: 72 % der Studienteilnehmer (Burosumab-Gruppe: 78 %, Placebo: 65 %) wiesen einen *BPI Schlimmsten Schmerz* von über 6 auf einer Skala von 0 – 10 auf. Zu Woche 12 zeigte sich jedoch bereits anhand der adjustierten mittleren Veränderung eine Tendenz zu einer stärkeren Abnahme des *BPI Schlimmsten Schmerzes* in der Burosumab-Gruppe (Burosumab: LsMean -0,63, Placebo: LsMean -0,38). Diese Tendenz setzte sich ebenfalls zu Woche 24 weiter fort. Die Behandlungsdifferenz zwischen den Behandlungsgruppen betrug -0,46, war jedoch nicht statistisch signifikant ( $p = 0,0919$ ) (Tabelle 4-45, Abbildung 4).

#### *Sensitivitätsanalyse aufgrund der fehlerhaften Stratifizierung zu Studienbeginn:*

Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 *Brief Pain Inventory* (BPI) *Schlimmster Schmerz* des patientenberichteten Fragebogens BPI und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der

Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens *BPI* bestimmt. Die verkürzte Form des Fragebogens enthielt die Fragen 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-46). Das für die Stratifizierung verwendete *Interactive web response system* (IWRS) war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*) verwendet wurde (Anhang 4-E 8b).

Tabelle 4-46: Nummerierung der einzelnen Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	Brief Pain Inventory
1	.. <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	.. <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	.. <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	.. <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.
9 A – G	.. <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:
a: IWRS: <i>Interactive web response system</i>		
b: für die Stratifizierung wurde eine verkürzte Form des BPI-Fragebogens verwendet, der nur die Items der Domäne Schmerzintensität enthielt.		

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI*

*Durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist.

Um einen möglichen Einfluss der fehlerhaften Stratifizierung auf den sekundären Wirksamkeitsendpunkt *BPI Schlimmster Schmerz* zu analysieren, wurde eine Sensitivitätsanalyse durchgeführt, bei der in das GEE-Modell anstelle der Kovariate *BPI Schlimmster Schmerz* die Kovariate *BPI Durchschnittlicher Schmerz* mit aufgenommen wurde. Die Ergebnisse waren übereinstimmend mit der Primäranalyse und bestätigen, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Ergebnisse der Studie UX023-CL303 hat. Da keine Abweichungen feststellbar waren, wird diese Analyse nicht im Abschnitt 4.3.1.3.1.4, sondern ausschließlich im Anhang 4-G dargestellt.

*Sensitivitätsanalysen aufgrund fehlender Post-Baseline Werte:*

Um den Einfluss von fehlenden Post-Baseline Werten auf die Ergebnisse zu überprüfen, wurde für den Endpunkt *BPI Schlimmster Schmerz* Sensitivitätsanalysen (LOCF, mBOCF, BOCF) durchgeführt. Die Ergebnisse waren übereinstimmend mit der Primäranalyse, wodurch bestätigt wird, dass eine Imputation von Werten keinen Einfluss auf die Aussagekraft der Ergebnisse hat. Daher werden diese Auswertungen nicht im Abschnitt 4.3.1.3.1.4, sondern ausschließlich im Anhang 4-G dargestellt.

Tabelle 4-47: Ergebnisse für *Schmerzerhebung – Responderanalyse für BPI Schlimmster Schmerz* aus RCT mit dem zu bewertenden Arzneimittel

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p- Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p- Wert <sup>d</sup>
<b>Anteil der Patienten mit einer Abnahme des BPI Schlimmsten Schmerzes um <math>\geq 15</math> %</b>						
Burosumab (68)	29 (42,6)	1,370 [0,668; 2,808]	0,3873	1,224 [0,796; 1,881]	0,078 [-0,086; 0,242]	0,3800
Placebo (66)	23 (34,8)					
<b>Anteil der Patienten mit einer Abnahme des BPI Schlimmsten Schmerzes um <math>\geq 30</math> %</b>						
Burosumab (68)	17 (25,0)	1,466 [0,625; 3,437]	0,3759	1,375 [0,713; 2,652]	0,068 [-0,071; 0,207]	0,4036
Placebo (66)	12 (18,2)					
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Responder; %: Anteil von N; [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>; RR: <i>Risk Ratio</i> (relatives Risiko), ARD: <i>Absolute Risk Difference</i> (absolute Risikodifferenz)</p> <p>Responder sind definiert als Studienteilnehmer, die nach 24 Wochen eine Abnahme des <i>BPI Schlimmsten Schmerz</i> um 15 % bzw. 30 % aufweisen.</p> <p>Ein RR bzw. OR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &lt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe, den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) und die Altersgruppen als unabhängige Variablen.</p> <p>b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>						

*Responderanalyse des BPI Schlimmsten Schmerzes*

Bei 42,6 % der Studienteilnehmer der Burosumab-Gruppe und bei 34,8 % der Placebo-Gruppe konnte im Verlauf bis zu Woche 24 eine Abnahme des chronischen Schmerzes um mindestens 15 % festgestellt werden (RR [95% KI]: 1,224 [0,796; 1,881], p-Wert = 0,3800) (Tabelle 4-47). Eine Abnahme des Endpunkts *BPI Schlimmster Schmerz* um 30 % erreichten dagegen 25 % der Studienteilnehmer der Burosumab-Gruppe und 18,2 % der Placebo-Gruppe (RR [95% KI]: 1,375 [0,713; 2,652], p-Wert = 0,4036) (Tabelle 4-47).

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-48: Ergebnisse für Schmerzhebung (BPI Schmerzintensität und BPI Schmerzinterferenz) aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	BPI (Brief Pain Inventory)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b>BPI Schmerzintensität</b>								
Placebo (N=66)	66 4,92 (1,547)	66 4,92 (1,547)	66 4,45 (1,897)	65 4,88 (1,528)	65 4,65 (2,130)	66 -0,47 (0,159) -0,38 (0,155)	65 -0,23 (0,210) -0,17 (0,196)	-0,43 [-0,93; 0,07] 0,0940
Burosumab (N=68)	68 5,18 (1,531)	68 5,18 (1,531)	68 4,56 (1,733)	67 5,17 (1,541)	67 4,45 (1,875)	68 -0,63 (0,148) -0,50 (0,157)	67 -0,73 (0,162) -0,60 (0,164)	-
<b>BPI Schmerzinterferenz</b>								
Placebo (N=66)	66 4,76 (2,174)	66 4,76 (2,174)	66 4,17 (2,442)	65 4,70 (2,138)	65 4,16 (2,385)	66 -0,59 (0,198) -0,29 (0,228)	65 -0,54 (0,212) -0,28 (0,258)	-0,13 [-0,70; 0,44] 0,6527
Burosumab (N=68)	68 5,23 (2,237)	68 5,23 (2,237)	68 4,30 (2,379)	67 5,18 (2,212)	67 4,36 (2,348)	68 -0,93 (0,198) -0,52 (0,213)	67 -0,81 (0,221) -0,41 (0,226)	-
MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> , Standardabweichung; SE: <i>standard error</i> (Standardfehler); H'g: Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i> a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 96. GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i> , berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline des Endpunkts <i>BPI Schmerzinterferenz/Schmerzintensität</i> als abhängige Variable und Region.								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention (N)	<i>BPI (Brief Pain Inventory)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>LsMean<sup>a</sup></b>
	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MW</b>	<b>MWD (SE)</b>	<b>MWD (SE)</b>	<b>[95% KI]</b>
	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>(SD)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>LsMean<sup>a</sup> (SE)</b>	<b>p-Wert</b>
								<b>falls p &lt; 0,05: Hedges'g [95% KI]</b>
Visite, Behandlungsarm, <i>BPI Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert des Endpunkts <i>BPI Schmerzinterferenz/Schmerzintensität</i> ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry</i> Matrix verwendet.								

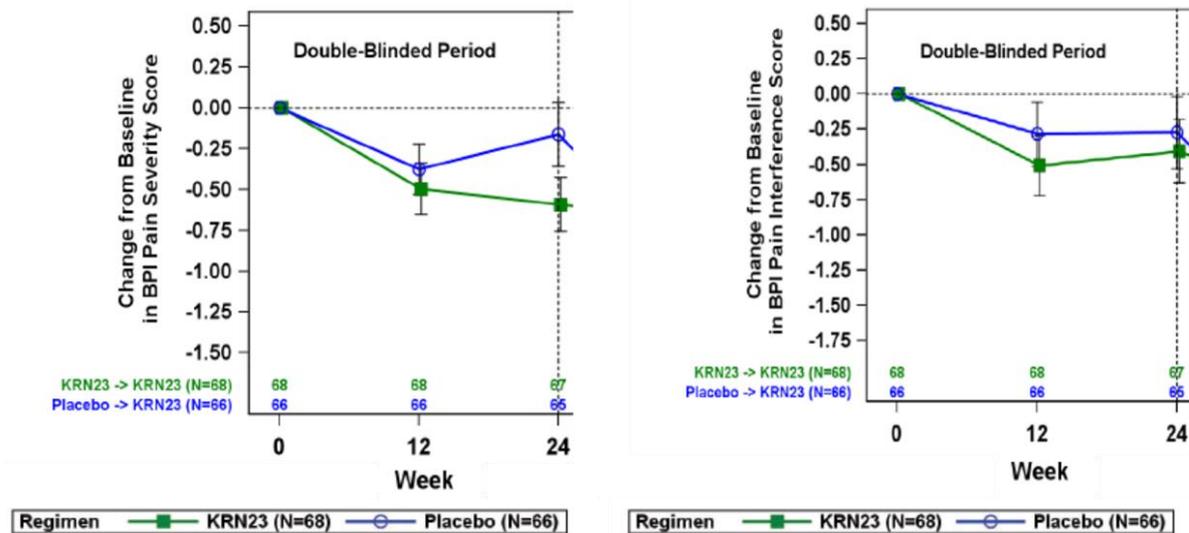


Abbildung 5: Mittelwerte für den Endpunkte *BPI Schmerzintensität* und *BPI Schmerzinterferenz* über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303

#### *BPI Schmerzintensität:*

Die Schmerzintensität wurde anhand von 4 Items mit einer 11-Punkte-Skala ermittelt. Zu Baseline wiesen die Burosumab- und die Placebo-Gruppe (MW: 5,18 vs. MW: 4,92) ähnliche Messwerte auf. Bereits zu Woche 12 zeigte sich jedoch eine Tendenz zu einer stärkeren Abnahme der Schmerzintensität durch Burosumab (LsMean -0,50) im Vergleich zu Placebo (LsMean -0,38). Diese Tendenz verstärkte sich weiterhin zu Woche 24 (Burosumab: LsMean -0,60, Placebo: LsMean -0,17). Die Behandlungsdifferenz von -0,43 erzielte jedoch mit  $p = 0,0940$  keine statistische Signifikanz (Tabelle 4-48, Abbildung 5).

#### *BPI Schmerzinterferenz:*

Zusätzlich zu dem *BPI Schlimmsten Schmerz* und der Domäne *BPI Schmerzintensität* wurde auch der Einfluss von Burosumab auf die *Schmerzinterferenz* evaluiert. Dies beinhaltet die Beeinträchtigung von XLH-Patienten bei Alltagsaktivitäten wie z. B. der Gehfähigkeit, dem Schlaf, der Arbeit und der Lebensqualität. Zu Baseline wies die Burosumab-Gruppe mit einem Mittelwert von 5,23 eine leicht höhere Schmerzinterferenz als die Placebo-Gruppe (MW: 4,76) auf. Nach 12 Wochen zeigte sich eine Tendenz in der Reduktion der Schmerzinterferenz zugunsten von Burosumab (LsMean: -0,52). In der Placebo-Gruppe betrug die adjustierte mittlere Änderung dagegen nur -0,29. Nach 24 Wochen war der Unterschied zu Baseline zwischen den beiden Behandlungsgruppen jedoch wieder leicht geringer (Burosumab: LsMean: -0,41, Placebo: LsMean: -0,28). Die Behandlungsdifferenz zwischen den beiden Gruppen von -0,13 war mit  $p = 0,6527$  nicht statistisch signifikant (Tabelle 4-48, Abbildung 5).

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse*

*durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte, kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie aufgrund der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

#### 4.3.1.3.1.5 Fatigue – RCT

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-49: Operationalisierung von *Fatigue*

Studie	Operationalisierung
UX023-CL303	<p>Der Fragebogen <i>Brief Fatigue Inventory (BFI)</i> ist ein patientenberichteter Fragebogen, der zwei Dimensionen, die Schwere der Fatigue (<i>BFI Schwere der Fatigue</i>) sowie den Einfluss der Fatigue auf das alltägliche Leben innerhalb der letzten 24 Stunden misst. Die Dimension <i>BFI Schwere der Fatigue</i> wird anhand von 3 Items abgefragt und dient der Messung des schwersten, gewöhnlichen und aktuellen Schweregrads, gemessen an jeweils einer Skala von 0 bis 10, wobei ein höherer Wert eine größere Einschränkung reflektiert. Der Einfluss der Fatigue auf das alltägliche Leben wird anhand von 6 Items (Aktivität, Stimmung, Gehfähigkeit, Arbeit, Beziehungen zu anderen und Lebensfreude) ebenfalls mit einer Skala von 0 bis 10 erfasst. Die Studienteilnehmer wurden jeweils 8 Tage vor der Studienvisite von einem Studienkoordinator telefonisch an das Ausfüllen des Fragebogens erinnert. Um das Ausmaß der Schwere der Fatigue (<i>BFI Schwere der Fatigue</i>) zu bestimmen, wurde von jeder Frage der Mittelwert aus den Messungen 7 Tage vor Studienvisite sowie dem Tag der Studienvisite gebildet. Der Score <i>BFI Globale Fatigue</i> (Frage 1 – 3 sowie Frage 4A – F) berechnet sich dagegen aus dem Mittelwert der 9 Items am jeweiligen Tag der Studienvisite. Der Fragebogen wurde sowohl zur Baseline-Visite als auch zu Woche 12 und Woche 24 erhoben.</p> <p><u>Analysepopulation:</u> FAS</p> <p><u>Statistisches Modell:</u></p> <ul style="list-style-type: none"> <li>• Auswertungen zur mittleren Änderung der <i>Endpunkte BFI Schlimmste Fatigue</i> bzw. <i>BFI Globale Fatigue</i> wurden zu Woche 12 und Woche 24 mit einem GEE-Modell bestimmt. Bei dem GEE-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline <i>BFI Schlimmste Fatigue</i> bzw. <i>BFI Globale Fatigue</i> als Kovariate. Die abhängige Variable war die Änderung zu Baseline im <i>Endpunkt BFI Schlimmste Fatigue</i> bzw. <i>BFI Globale Fatigue</i>. Als Kovarianzstruktur wurde eine <i>Compound Symmetry</i> Matrix verwendet.</li> </ul> <p><u>Darstellung im Dossier:</u></p> <ul style="list-style-type: none"> <li>• <i>BFI Schlimmste Fatigue</i> <ul style="list-style-type: none"> <li>○ Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen</li> </ul> </li> <li>• <i>BFI Globale Fatigue</i> <ul style="list-style-type: none"> <li>○ Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier mittels Hedges'g geprüft</li> </ul> </li> </ul>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-50: Bewertung des Verzerrungspotenzials für *Fatigue* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Der *BFI*-Fragebogen ist ein patientenberichteter Fragenbogen, der kein Training der Studienteilnehmer erforderte. Sowohl die Studienteilnehmer als auch das Studienpersonal und der Prüfarzt waren während der 24-wöchigen, doppelblinden, Placebo-kontrollierten Phase der Studie gegenüber der Behandlung verblindet. Da alle Studienteilnehmer während der 24-wöchigen Phase mindestens einmal die Prüfsubstanz erhielten, umfasst die FAS die gesamte Studienpopulation. Somit fand eine ergebnisunabhängige Berichterstattung unter adäquater Umsetzung des ITT-Prinzips statt.

Der Rücklauf des Fragebogens in der Studie belief sich zu allen Messzeitpunkten auf über 70 %.

Zusammenfassend liegt für den Endpunkt *Fatigue* (*BFI Schlimmste Fatigue*, *BFI Globale Fatigue*) in der randomisierten doppelblinden Placebo-kontrollierten Studie UX023-CL303 ein **niedriges Verzerrungspotenzial** vor.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-51: Ergebnisse für *Fatigue* aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	<i>BFI (Brief Fatigue Inventory)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b><i>BFI Schlimmste Fatigue</i></b>								
Placebo (N=66)	66 6,74 (1,526)	66 6,74 (1,526)	66 5,99 (1,862)	65 6,72 (1,533)	65 6,07 (2,001)	66 -0,75 (0,212) -0,53 (0,266)	65 -0,65 (0,239) -0,45 (0,298)	-0,20 [0,40; -0,80]
Burosumab (N=68)	68 6,94 (1,657)	68 6,94 (1,657)	68 6,22 (2,007)	67 6,92 (1,664)	67 5,99 (2,192)	68 -0,72 (0,194) -0,44 (0,261)	67 -0,93 (0,224) -0,65 (0,280)	0,5102 -
<b><i>BFI Globaler Fatigue Score</i></b>								
Placebo (N=66)	66 4,86 (1,932)	66 4,86 (1,932)	66 4,16 (2,315)	65 4,82 (1,915)	65 4,20 (2,341)	66 -0,70 (0,202) -0,14 (0,262)	65 -0,62 (0,215) -0,08 (0,292)	0,13 [0,70; -0,44]
Burosumab (N=68)	68 5,37 (2,044)	68 5,37 (2,044)	68 4,48 (2,342)	67 5,33 (2,022)	67 4,68 (2,306)	68 -0,90 (0,216) -0,18 (0,261)	67 -0,65 (0,221) 0,05 (0,261)	0,6572 -
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: <i>standard deviation</i> (Standardabweichung); SE: <i>standard error</i> (Standardfehler); H'g: Hedges'g (wurde <i>post-hoc</i> berechnet); LsMean: <i>Least Squares Means</i></p> <p>a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 96.</p> <p>GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i>, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline des entsprechenden <i>BFI</i> Endpunkts als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Der Baseline-Wert des entsprechenden <i>BFI</i> Endpunkts ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry Matrix</i> verwendet.</p>								

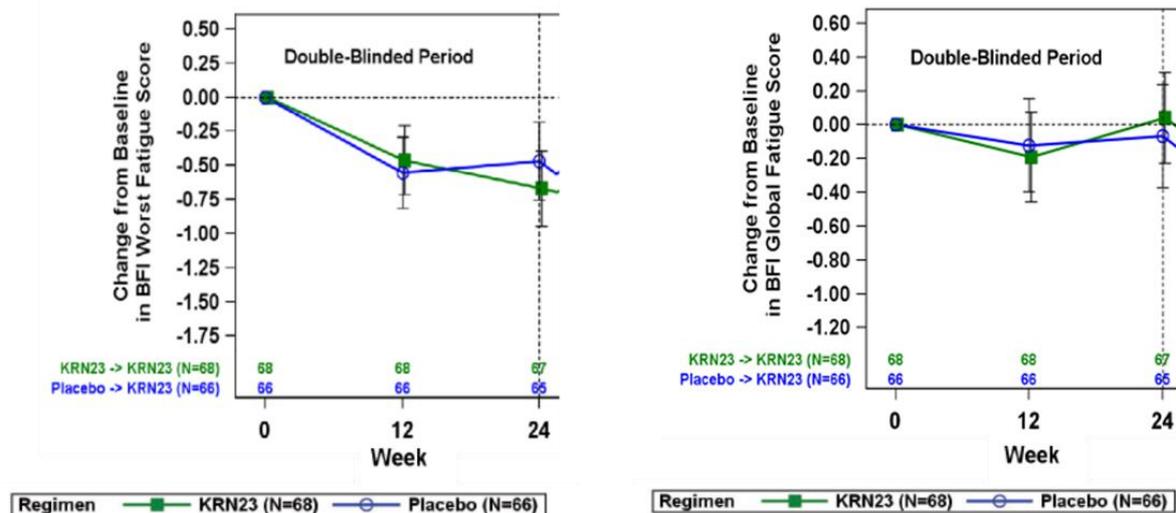


Abbildung 6: Mittelwerte für den Endpunkt Fatigue (*BFI Schlimmste Fatigue* und *BFI Globaler Fatigue*) über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303

Der chronische Phosphatverlust bei Erwachsenen mit XLH beeinträchtigt auf vielfältige Weise den Alltag der Patienten. Neben Knochenschmerzen führt der Verlust an Phosphat zu einer verringerten ATP-Bildung, reduzierter Muskelkraft und Fatigue. Um den Einfluss von Burosumab auf die Fatigue zu evaluieren, wurden sowohl die Schwere der Fatigue als auch der Einfluss von Fatigue auf das alltägliche Leben innerhalb der letzten 24 h anhand des *Brief Fatigue Inventory* gemessen. Die Schwere der Fatigue umfasst 3 Fragen, der Einfluss von Fatigue auf den Alltag insgesamt 9 Fragen. Die Messung erfolgte anhand einer Skala von jeweils 0 – 10, wobei ein höherer Wert eine stärkere Ausprägung der Fatigue bzw. größeren Einfluss auf Alltagsaktivitäten reflektiert.

#### *BFI Schlimmste Fatigue*

Die von den Studienteilnehmern verspürte *BFI Schlimmste Fatigue* innerhalb der letzten 24 Stunden war zwischen beiden Behandlungsgruppen zu Baseline vergleichbar (Burosumab: MW: 6,94, Placebo: MW: 6,74). Nach 12 Wochen war in beiden Gruppen eine leichte Abnahme der Fatigue feststellbar (Burosumab: LsMean: -0,44; Placebo: LsMean: -0,53). Zu Woche 24 betrug die adjustierte Änderung in der Burosumab-Gruppe -0,65 und in der Placebo-Gruppe -0,45. Die Behandlungsdifferenz von -0,20 zu Baseline erreichte jedoch keine statistische Signifikanz ( $p = 0,5102$ , GEE-Modell) (Tabelle 4-51, Abbildung 6).

#### *BFI Globale Fatigue*

Der *BFI Globale Fatigue* Score betrug zu Baseline in der Burosumab-Gruppe 5,37 und in der Placebo-Gruppe 4,86. Nach 12 Wochen konnte ebenfalls wie bei dem Endpunkt *BFI Schlimmste Fatigue* in beiden Gruppen eine leichte Abnahme der Fatigue verzeichnet werden (Burosumab: LsMean: -0,18; Placebo: LsMean: -0,14). Dieser Wert änderte sich zu Woche 24 nur gering. Die Behandlungsdifferenz betrug 0,13 und war mit  $p = 0,6572$  nicht statistisch signifikant (Tabelle 4-51, Abbildung 6).

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

#### 4.3.1.3.1.6 Erfassung von Therapieeffekten – RCT

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-52: Operationalisierung von Erfassung von Therapieeffekten

Studie	Operationalisierung
UX023-CL303	<p>Der <i>Patient Global Impression of Severity/Improvement (PGI-S, PGI-I)</i> ist ein patientenberichteter Fragebogen, der die subjektive Wahrnehmung eines Patienten bezüglich des Therapieerfolgs misst. In der Studie UX023-CL303 wurde die globale Verbesserung des Gesundheitszustandes mit dem <i>PGI-I</i> anhand einer kategorischen 7-Punkte-Skala, die von sehr stark verbessert bis sehr stark verschlechtert reicht, ermittelt:</p> <p>Die subjektive Wahrnehmung der Krankheitsschwere zu Baseline (<i>PGI-S</i>) wurde dagegen mit einer kategorischen 4-Punkte-Skala ermittelt, die von 1 (normal) bis 4 (schwer) reicht: Der <i>PGI-S</i> wurde ausschließlich zu Baseline erhoben, der <i>PGI-I</i> dagegen zu Woche 12 und Woche 24.</p> <p><u>Analysepopulation:</u> FAS</p> <p><u>Statistisches Modell:</u></p> <ul style="list-style-type: none"> <li>• Auswertungen zur mittleren Änderung PGI wurden zu Woche 12 und Woche 24 mit einem GEE-Modell bestimmt. Bei dem GEE-Modell wurden als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt. Die abhängige Variable war der <i>PGI</i>. Als Kovarianzstruktur wurde eine <i>Compound Symmetry</i> Matrix verwendet.</li> </ul> <p><u>Darstellung im Dossier:</u></p> <ul style="list-style-type: none"> <li>• Differenz der mittleren Änderung von Baseline zu Woche 24 zwischen den Behandlungsgruppen</li> <li>• Klinische Relevanz: Die klinische Relevanz der beobachteten Veränderung wird im vorliegenden Dossier mittels Hedges'g geprüft</li> </ul>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-53: Bewertung des Verzerrungspotenzials für *Erfassung von Therapieeffekten* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

*Begründen Sie für jede Studie die abschließende Einschätzung.*

Der *PGI-I*-Fragebogen ist ein patientenberichteter Fragenbogen, der kein Training der Studienteilnehmer erforderte. Sowohl die Studienteilnehmer als auch das Studienpersonal und der Prüfarzt waren während der 24-wöchigen, doppelblinden, Placebo-kontrollierten Phase der Studie gegenüber der Behandlung verblindet. Da alle Studienteilnehmer während der 24-wöchigen Phase mindestens einmal die Prüfsubstanz erhielten, umfasst die FAS die gesamte Studienpopulation. Somit fand eine ergebnisunabhängige Berichterstattung unter adäquater Umsetzung des ITT-Prinzips statt.

Der Rücklauf des Fragebogens in der Studie belief sich zu allen Messzeitpunkten auf über 70 %.

Zusammenfassend liegt für den Endpunkt *Erfassung von Therapieeffekten* in der randomisierten doppelblinden Placebo-kontrollierten Studie UX023-CL303 ein **niedriges Verzerrungspotenzial** vor.

*Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.*

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-54: Ergebnisse für Erfassung von Therapieeffekten aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	PGI-I							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
Placebo (N=66)	66 3,00 (0,823)	66 3,00 (0,823)	66 3,61 (1,122)	64 3,00 (0,816)	64 3,84 (1,394)	_ <sup>b</sup> _ <sup>b</sup> 3,77 (0,179)	_ <sup>b</sup> _ <sup>b</sup> 3,99 (0,211)	-0,29 [-0,74; 0,16] 0,2073
Burosumab (N=68)	68 3,18 (0,622)	67 3,16 (0,618)	67 3,49 (1,160)	67 3,16 (0,618)	67 3,54 (1,235)	_ <sup>b</sup> _ <sup>b</sup> 3,65 (0,165)	_ <sup>b</sup> _ <sup>b</sup> 3,70 (0,171)	-

MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SD: *standard deviation* (Standardabweichung); SE: *standard error* (Standardfehler); H'g: Hedges'g (wurde *post-hoc* berechnet); LsMean: *Least Squares Means*

a: Berechnung mit *generalized estimation equation* (GEE)-Modell zu Woche 96.

b: Nicht bestimmt.

GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts *BPI Durchschnittlichen Schmerz*, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit *PGI* als abhängige Variable und Region, Visite, Behandlungsarm, *BPI Durchschnittlichen Schmerz* (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Als Kovarianzstruktur wird eine *Compound Symmetry* Matrix verwendet.

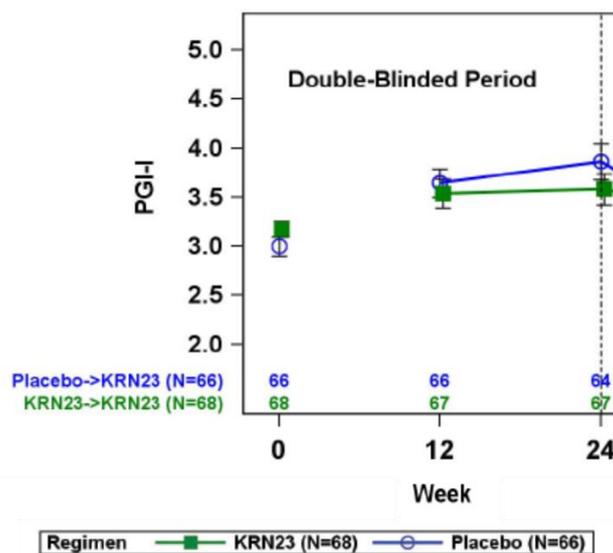


Abbildung 7: Mittelwerte für den Endpunkt *Erfassung von Therapieeffekten* über den Verlauf der 24-wöchigen doppelblinden Phase der Studie UX023-CL303.

Die vom Patienten wahrgenommene Verbesserung des Gesundheitszustandes wurde anhand des *PGI-I* Fragebogens erhoben. Die kategorische Skala reicht von 1-7, wobei 7 als sehr stark verschlechtert und 1 als stark verbessert definiert ist.

Zu Beginn der Studie wurde der Gesundheitszustand in der Burosumab- und Placebo-Gruppe als ähnlich beeinträchtigt wahrgenommen (Burosumab: MW: 3,18; Placebo: MW: 3,0). Nach 12 Wochen unterschieden sich die beiden Behandlungsgruppen kaum voneinander (Burosumab: LsMean: 3,49; Placebo: LsMean: 3,61). Auch nach 24 Wochen war keine deutliche Änderung des vom Patienten wahrgenommenen Gesundheitszustandes zwischen der Burosumab- und Placebo-Gruppe zu verzeichnen. Die Änderung zu Baseline betrug -0,29 und war mit  $p = 0,2073$  nicht statistisch signifikant (Tabelle 4-54, Abbildung 7).

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren)

rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

**4.3.1.3.1.7 Sicherheit – RCT**

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-55: Operationalisierung von Sicherheit

Studie	Operationalisierung
UX023-CL303	<p>Alle <i>unerwünschten Ereignisse (UE)</i> wurden gemäß MedDRA ab dem Tag der Unterzeichnung der Einwilligungserklärung bis einschließlich 12 Tage nach der letzten Gabe der Prüfsubstanz aufgezeichnet. In die Auswertung wurden jedoch nur <i>UE</i> eingeschlossen, die nach der erstmaligen Verabreichung der Prüfsubstanz (Burosumab) oder Placebo auftraten (<i>Treatment emergent adverse events (TEAE)</i>), um die Inzidenz, Häufigkeit und Schwere der <i>UE</i> sowie einen möglichen Zusammenhang mit der Behandlung zu bewerten. <i>UE</i>, die zwischen dem Zeitpunkt der Einwilligung und der Baseline-Visite auftraten, wurden hierbei nicht berücksichtigt, da diese <i>UE</i> in keinem Bezug zur Behandlung mit der Prüfsubstanz standen und die objektive Beurteilung des Sicherheitsprofils in beiden Behandlungsarmen beeinträchtigt hätten.</p> <p>Bei jeder Studienvisite oder Hausvisite wurden die Studienteilnehmer nach den seit der letzten Visite neu aufgetretenen oder andauernden <i>UE</i> befragt und fanden sowohl in der Placebo- als auch in der Burosumab-Gruppe statt. Zeitgleich wurden die Begleitmedikationen sowie zusätzliche Therapien erfasst. Hierbei wurde besonders darauf geachtet, dass die Dosis der Schmerzmedikation über die gesamte Studiendauer konstant gehalten wurde.</p> <p>Folgende Klassifizierungen der <i>TEAE</i> wurden vorgenommen:</p> <p><u>Schweregrad der <i>TEAE</i>:</u></p> <p>Die Einteilung der <i>TEAE</i> erfolgte anhand von <i>NCI Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.03</i>. Für den Fall, dass eine Einstufung anhand dieser Kriterien nicht möglich war, wurden die <i>TEAE</i> anhand folgender Kriterien als mild, moderat, schwer, lebensbedrohlich oder tödlich eingestuft.</p> <ul style="list-style-type: none"> <li>• Mild (Grad 1): Hierzu zählen alle Anzeichen von Symptomen, die leicht tolerierbar sind und nur geringe Reizungen verursachen und zu keiner Einschränkung in der allgemeinen Aktivität des Patienten führen. Zudem ist keine ärztliche Behandlung notwendig, da diese nur vorübergehend auftreten.</li> <li>• Moderat (Grad 2): <i>TEAE</i>, die in einem geringen Maß zu Unwohlsein oder Besorgnis führen und den Patienten im Alltag möglicherweise einschränken und die körperliche Funktion beeinträchtigen, aber durch einfache therapeutische Maßnahmen verbessert werden können.</li> <li>• Schwer (Grad 3): <i>TEAE</i>, die zu einer Unterbrechung der Alltagsaktivitäten sowie Beeinträchtigung der Studienteilnehmer führen und eine systemische Therapie oder anderweitige Behandlung erfordern.</li> <li>• Lebensbedrohlich (Grad 4): <i>TEAE</i>, bei denen der Studienteilnehmer dem Risiko ausgesetzt wird zu sterben oder UEs, die zu Behinderungen führen.</li> <li>• Tod (Grad 5): <i>TEAE</i>, die zum Tod von Studienteilnehmern führen.</li> </ul> <p><u>Kausalität der <i>TEAE</i>:</u></p> <p>Anhand folgender Kategorien wurde vom Prüfarzt der Zusammenhang zwischen unerwünschten Ereignissen und der Prüfsubstanz ermittelt:</p> <ul style="list-style-type: none"> <li>• Nichtzusammenhängend: <i>TEAE</i>, die ohne jeglichen Zweifel definitiv nicht mit der Studienmedikation oder dem Verfahren im Zusammenhang stehen.</li> <li>• Wahrscheinlich nicht zusammenhängend: <i>TEAE</i>, deren Zusammenhang mit der Prüfsubstanz oder dem Verfahren eher zweifelhaft ist.</li> </ul>

Studie	Operationalisierung
	<ul style="list-style-type: none"> <li>• Möglicherweise zusammenhängend: <i>TEAE</i>, die möglicherweise mit der Prüfsubstanz oder dem Verfahren in Zusammenhang stehen.</li> <li>• Mutmaßlich zusammenhängend: <i>TEAE</i>, bei denen ein Zusammenhang mit der Verabreichung der Prüfsubstanz oder dem Verfahren möglich ist.</li> <li>• Definitiv zusammenhängend: <i>TEAE</i>, die eindeutig mit der Prüfsubstanz oder dem Verfahren assoziiert sind.</li> </ul> <p><u>Schwerwiegende <i>TEAE</i>:</u> Zu den schwerwiegenden <i>TEAE</i> zählen folgende Ereignisse:</p> <ul style="list-style-type: none"> <li>• Tod</li> <li>• Lebensbedrohliches <i>TEAE</i></li> <li>• <i>TEAE</i>, die eine Hospitalisierung erfordern oder einen Krankenhausaufenthalt verlängern</li> <li>• Persistierende oder signifikante Einschränkung oder Behinderung (wesentliche Beeinträchtigung der normalen Lebensfunktion)</li> <li>• kongenitale Anomalie/Geburtsfehler</li> </ul> <p>Hospitalisierungen, die bereits vor Studienbeginn geplant waren, wurden nicht als schweres <i>UE</i> eingestuft. Dagegen wurden Hospitalisierungen, die, nachdem die Rekrutierungsphase bereits abgeschlossen war, aber aufgrund von Vorerkrankungen notwendig waren, als schweres <i>UE</i> eingestuft.</p> <p>Wichtige medizinische Ereignisse, die nicht tödlich oder direkt lebensbedrohlich verliefen, oder eine Hospitalisierung zur Folge hatten, wurden als schweres <i>UE</i> angesehen, sobald, basierend auf medizinischen Beurteilungen, diese den Patienten gefährdeten, oder eine medizinische Maßnahme oder Operation benötigt wurde, um die oben aufgeführten schweren <i>UEs</i> zu vermeiden.</p> <p><u>Ein <i>TEAE</i> von besonderem Interesse wird definiert als <i>TEAE</i> der folgenden Kategorien:</u></p> <ul style="list-style-type: none"> <li>• <i>Reaktionen an der Einstichstelle</i></li> <li>• <i>Hypersensitivität</i></li> <li>• <i>Hyperphosphatämie</i></li> <li>• <i>Ektopische Mineralisierung</i></li> <li>• <i>Restless-Legs-Syndrom</i></li> </ul> <p><i>SOC und PT nach MedDRA</i></p> <p>Die <i>TEAE</i> wurden gemäß MedDRA kodiert und entsprechend den SOC und PT klassifiziert. Im Dossier werden auf Ebene der SOC und PT folgende Auswertungen dargestellt:</p> <ul style="list-style-type: none"> <li>• Jegliche <i>TEAE</i>: Alle SOC und PT, die von mind. <math>\geq 10</math> % der Patienten in einem Studienarm berichtet wurden</li> <li>• Schwerwiegende <i>TEAE</i>: Alle SOC und PT, die von mind. <math>\geq 5</math> % der Patienten in einem Studienarm berichtet wurden</li> <li>• <i>TEAE</i> mit NCI-CTCAE-Schweregrad 3 – 4: Alle SOC und PT, die von mind. <math>\geq 5</math> % der Patienten in einem Studienarm berichtet wurden</li> </ul> <p><u>Analysepopulation: <i>Safety</i></u></p> <p><u>Statistisches Modell:</u></p> <p>Im Studienprotokoll der Studie UX023-CL303 war prospektiv eine deskriptive Darstellung des Anteils der Patienten mit <i>TEAE</i> geplant. <i>Post-hoc</i> wurden für das vorliegende Dossier sowohl für die Gesamtraten als auch auf Ebene von SOC und PT Effektschätzer berechnet,</p>

Studie	Operationalisierung
	<p>um die Anforderungen der Dossievorlage zu erfüllen.</p> <ul style="list-style-type: none"> <li>• Das OR wurde mittels der logistischen Regression inklusive 95% KI und p-Wert bestimmt. Die Behandlungsgruppe und der tatsächliche Stratifizierungsfaktor (<i>BPI durchschnittlicher Schmerz</i>) wurden als unabhängige Variablen in das Modell aufgenommen.</li> <li>• Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation und der p-Wert mittels Fishers exaktem Test berechnet.</li> </ul> <p><u>Erkrankungsbezogene Ereignisse:</u></p> <p>Laut der aktuellen Dossievorlage sollen bei der Erhebung <i>UE</i> zusätzliche <i>UE</i>-Analysen durchgeführt werden, bei denen krankheitsbezogene Ereignisse unberücksichtigt bleiben. In der Nutzenbewertung vom 01.02.2021 wurde bereits darauf hingewiesen, dass zwar <i>UE</i> der Systemorganklasse Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen sowie die PT Zahnabszess und Karies typische Symptome der betrachteten Grunderkrankung darstellen können, einzelne Symptome wie z. B. Schmerzen in den Extremitäten sich aber nicht nachträglich aus den betreffenden Systemorganklassen separieren lassen [37]. Aufgrund dieser Problematik werden im Folgenden ausschließlich die Gesamtanalysen der jeglichen, schweren und schwerwiegenden <i>UE</i> dargestellt, bei denen nicht ausgeschlossen werden kann, dass krankheitsbezogene Ereignisse beinhaltet sind.</p> <p><u>Darstellung im Dossier:</u></p> <ul style="list-style-type: none"> <li>• Gesamtrate <i>TEAE</i> bis Woche 24</li> <li>• Gesamtrate <i>schwerer TEAE</i> bis Woche 24</li> <li>• Gesamtrate <i>schwerwiegender TEAE</i> bis Woche 24</li> <li>• <i>TEAE</i> unterteilt nach MedDRA SOC und PT, die bei mind. <math>\geq 10\%</math> der Studienteilnehmer in mindestens einer Behandlungsgruppe aufgetreten sind</li> <li>• <i>Schwere TEAE</i> unterteilt nach MedDRA SOC und PT, die bei mind. <math>\geq 5\%</math> der Studienteilnehmer in mindestens einer Behandlungsgruppe aufgetreten sind</li> <li>• <i>Schwerwiegende TEAE</i> unterteilt nach MedDRA SOC und PT, die bei mind. <math>\geq 5\%</math> der Studienteilnehmer in mindestens einer Behandlungsgruppe aufgetreten sind</li> <li>• <i>TEAE von besonderem Interesse (unabhängig von der Häufigkeit)</i></li> </ul>

*Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.*

*Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.*

Tabelle 4-56: Bewertung des Verzerrungspotenzials für *Sicherheit* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
UX023-CL303	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte doppelblinde Placebo-kontrollierte multizentrische Studie der Phase III.

Für die Analysen wurde die *Safety*-Population herangezogen. Hierbei handelt es sich um alle Studienteilnehmer, die mindestens eine Dosis der Studienmedikation erhalten haben.

Zusammenfassend liegt für den Endpunkt *Sicherheit* ein **niedriges Verzerrungspotenzial** vor.

Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

**Unerwünschte Ereignisse nach Beginn der Behandlung (TEAE)**Tabelle 4-57: Ergebnisse für *unerwünschte Ereignisse nach Beginn der Behandlung (TEAE)* aus RCT mit dem zu bewertenden Arzneimittel (*Safety*-Population)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p- Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p- Wert <sup>d</sup>
<b>Gesamtrate - Baseline zu Woche 24</b>						
<b>Jegliche TEAE</b>						
Burosumab (68)	64 (94,1)	1,206 [0,301; 4,832]	0,7897	1,018 [0,930; 1,115]	0,017 [-0,068; 0,102]	0,7425
Placebo (66)	61 (92,4)					
<b>Schwerwiegende TEAE</b>						
Burosumab (68)	2 (2,9)	1,697 [0,132; 21,788]	0,6827	1,941 [0,180; 20,900]	0,014 [-0,036; 0,064]	1,0000
Placebo (66)	1 (1,5)					
<b>TEAE Grad 3 – 4</b>						
Burosumab (68)	8 (11,8)	0,956 [0,331; 2,761]	0,9338	0,971 [0,387; 2,435]	-0,004 [-0,113; 0,106]	1,0000
Placebo (66)	8 (12,1)					
<b>Studienabbruch durch TEAE</b>						
Burosumab (68)	0 (0)	-	-	-	-	-
Placebo (66)	0 (0)					
<b>Therapieabbruch durch TEAE</b>						
Burosumab (68)	0 (0)	-	-	-	-	-
Placebo (66)	0 (0)					
<b>Tod durch TEAE</b>						
Burosumab (68)	0 (0)	-	-	-	-	-
Placebo (66)	0 (0)					
<p>N: Patientenzahl der entsprechend definierten Population; n: Anzahl der Patienten mit mindestens einem Ereignis in der entsprechenden Kategorie; %: Anteil von N; <i>Treatment emergent adverse event</i> (unerwünschtes Ereignis nach Beginn der Behandlung); [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>; RR: <i>Risk Ratio</i> (relatives Risiko); ARD: <i>Absolute Risk Difference</i> (Absolute Risikodifferenz)</p> <p>Ein RR bzw. OR &lt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &gt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe und den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) als unabhängige Variablen.</p> <p>b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>						

Während der doppelblinden 24-wöchigen Phase hatten die meisten Studienteilnehmer ein TEAE unabhängig von der Behandlung (Burosumab: 64 (94,1 %) und Placebo: 61 (92,4 %)).

Schwere *TEAE* traten lediglich bei 2 (2,9 %) Studienteilnehmer in der Burosumab-Gruppe und bei 1 (1,5 %) Studienteilnehmer in der Placebo-Gruppe auf. *TEAE* von Grad 3 waren in der Burosumab-Gruppe (8 (11,8 %)) und in der Placebo-Gruppe (8 (12,1 %)) gleich häufig vorhanden. *TEAE* des Grads 4 waren während der 24-wöchigen Studiendauer nicht zu beobachten. Die restlichen *TEAE* waren mild oder moderat (Grad 1 und 2). Es konnten keine statistisch signifikanten Unterschiede zwischen den beiden Behandlungsgruppen festgestellt werden (Tabelle 4-57).

***TEAE nach MedDRA SOC und PT (bei  $\geq 10$  % der Patienten in mindestens einer Behandlungsgruppe)***

Tabelle 4-58: Ergebnisse für *TEAE nach System Organ Class (SOC) und Preferred Term (PT)* aus RCT mit dem zu bewertenden Arzneimittel (*Safety-Population*)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p- Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p- Wert <sup>d</sup>
<b>Baseline zu Woche 24</b>						
<b>SOC Skelettmuskulatur-, Bindegewebs-und Knochenkrankungen</b>						
Burosumab (68)	26 (38,2)	0,759 [0,372; 1,549]	0,4452	0,841 [0,563; 1,257]	0,072 [-0,094; 0,239]	0,4838
Placebo (66)	30 (45,5)					
<b>PT Arthralgie</b>						
Burosumab (68)	6 (8,8)	0,299 [0,106; 0,842]	0,0227	0,364 [0,152; 0,873]	0,154 [0,031; 0,278]	0,0198
Placebo (66)	16 (24,2)					
<b>PT Rückenschmerzen</b>						
Burosumab (68)	10 (14,7)	1,796 [0,596; 5,409]	0,2954	1,618 [0,623; 4,199]	0,056 [-0,053; 0,165]	0,4260
Placebo (66)	6 (9,1)					
<b>PT Schmerz in einer Extremität</b>						
Burosumab (68)	6 (8,8)	0,556 [0,186; 1,665]	0,2915	0,582 [0,224; 1,511]	0,063 [-0,046; 0,173]	0,2961
Placebo (66)	10 (15,2)					
<b>SOC Infektionen und parasitäre Erkrankungen</b>						
Burosumab (68)	34 (50,0)	1,148 [0,577; 2,287]	0,6919	1,065 [0,750; 1,510]	0,030 [-0,139; 0,199]	0,7333
Placebo (66)	31 (47,0)					
<b>PT Zahnabszesse</b>						
Burosumab (68)	9 (13,2)	1,581 [0,513; 4,870]	0,4220	1,456 [0,549; 3,863]	0,041 [-0,065; 0,148]	0,5858
Placebo (66)	6 (9,1)					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

<b>UX023-CL303 Population (N)</b>	<b>Patienten n (%)</b>	<b>OR<sup>a,b</sup> [95% KI]</b>	<b>p- Wert<sup>a</sup></b>	<b>RR<sup>b,c</sup> [95% KI]</b>	<b>ARD<sup>b,c</sup> [95% KI]</b>	<b>p- Wert<sup>d</sup></b>
<b><i>PT Nasopharyngitis</i></b>						
Burosumab (68)	11 (16,2)	1,957 [0,668; 5,731]	0,2185	1,779 [0,698; 4,534]	0,071 [-0,041; 0,183]	0,3002
Placebo (66)	6 (9,1)					
<b>SOC Erkrankungen des Nervensystems</b>						
Burosumab (68)	26 (38,2)	1,926 [0,895; 4,143]	0,0931	1,484 [0,892; 2,469]	0,125 [-0,032; 0,281]	0,1411
Placebo (66)	17 (25,8)					
<b><i>PT Kopfschmerzen</i></b>						
Burosumab (68)	9 (13,2)	1,889 [0,588; 6,069]	0,2832	1,747 [0,618; 4,940]	0,057 [-0,046; 0,159]	0,3986
Placebo (66)	5 (7,6)					
<b><i>PT Restless-Legs-Syndrom</i></b>						
Burosumab (68)	8 (11,8)	1,687 [0,510; 5,577]	0,3886	1,553 [0,535; 4,504]	0,042 [-0,058; 0,142]	0,5616
Placebo (66)	5 (7,6)					
<b><i>PT Schwindelgefühl</i></b>						
Burosumab (68)	8 (11,8)	2,251 [0,620; 8,166]	0,2152	1,941 [0,614; 6,140]	0,057 [-0,039; 0,153]	0,3657
Placebo (66)	4 (6,1)					
<b>SOC Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>						
Burosumab (68)	22 (32,4)	1,197 [0,564; 2,539]	0,6378	1,124 [0,674; 1,875]	0,036 [-0,120; 0,192]	0,7098
Placebo (66)	19 (28,8)					
<b><i>PT Fatigue</i></b>						
Burosumab (68)	6 (8,8)	0,846 [0,263; 2,724]	0,7773	0,832 [0,295; 2,345]	0,018 [-0,082; 0,118]	0,7772
Placebo (66)	7 (10,6)					
<b>SOC Erkrankungen des Gastrointestinaltrakts</b>						
Burosumab (68)	21 (30,9)	1,213 [0,563; 2,617]	0,6195	1,132 [0,666; 1,925]	0,036 [-0,118; 0,190]	0,7057
Placebo (66)	18 (27,3)					
<b><i>PT Übelkeit</i></b>						
Burosumab (68)	7 (10,3)	1,209 [0,373; 3,918]	0,7500	1,132 [0,402; 3,192]	0,012 [-0,088; 0,112]	1,0000
Placebo (66)	6 (9,1)					
<b>SOC Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>						
Burosumab (68)	12 (17,6)	0,808 [0,338; 1,934]	0,6297	0,832 [0,416; 1,663]	0,036 [-0,098; 0,170]	0,6654
Placebo (66)	14 (21,2)					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p- Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p- Wert <sup>d</sup>
<b>SOC Untersuchungen</b>						
Burosumab (68)	14 (20,6)	1,706 [0,671; 4,340]	0,2597	1,510 [0,702; 3,247]	0,070 [-0,057; 0,196]	0,3612
Placebo (66)	9 (13,6)					
<b>SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>						
Burosumab (68)	11 (16,2)	0,980 [0,386; 2,491]	0,9665	0,971 [0,452; 2,084]	0,005 [-0,121; 0,130]	1,0000
Placebo (66)	11 (16,7)					
<b>PT Schmerzen im Oropharynx</b>						
Burosumab (68)	1 (1,5)	0,116 [0,013; 1,015]	0,0516	0,139 [0,018; 1,096]	0,091 [0,012; 0,171]	0,0317
Placebo (66)	7 (10,6)					
<b>SOC Erkrankungen der Haut und des Unterhautzellgewebes</b>						
Burosumab (68)	10 (14,7)	1,771 [0,595; 5,267]	0,3017	1,618 [0,623; 4,199]	0,056 [-0,053; 0,165]	0,4260
Placebo (66)	6 (9,1)					
<b>SOC Stoffwechsel- und Ernährungsstörungen</b>						
Burosumab (68)	9 (13,2)	1,552 [0,512; 4,709]	0,4345	1,456 [0,549; 3,863]	0,041 [-0,065; 0,148]	0,5858
Placebo (66)	6 (9,1)					
<p>N: Patientenzahl der entsprechend definierten Population; n: Anzahl der Patienten mit mindestens einem Ereignis in der entsprechenden Kategorie; %: Anteil von N; <i>Treatment emergent adverse event</i> (unerwünschtes Ereignis nach Beginn der Behandlung); SOC: <i>System Organ Class</i> (Systemorganklasse); PT: <i>Preferred Term</i>, [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>; RR: <i>Risk Ratio</i> (relatives Risiko); ARD: <i>Absolute Risk Difference</i> (Absolute Risikodifferenz)</p> <p>Ein RR bzw. OR &lt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &gt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe und den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) als unabhängige Variablen.</p> <p>b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>						

Für fast alle *TEAE* auf Ebene der MedDRA SOC und PT (die bei  $\geq 10$  % der Patienten in mindestens einer Behandlungsgruppe aufgetreten sind) liegen keine statistisch signifikanten Unterschiede vor (Tabelle 4-58).

Es zeigte sich jedoch in der *SOC Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen* ein statistisch signifikanter Unterschied zugunsten von Burosumab hinsichtlich des *PT Arthralgie* zu Woche 24 (RR [95% KI]: 0,364 [0,152; 0,873]; p = 0,0198) (Tabelle 4-58). Des Weiteren zeigte sich in der *SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums* hinsichtlich des *PT Schmerzen im Oropharynx* ein signifikanter Unterschied zu Woche 24 zugunsten von Burosumab (RR [95% KI]: 0,139 [0,018; 1,096]; p = 0,0317) (Tabelle 4-58).

*Schwerwiegende TEAE* und *TEAE* des Grad 3 und 4 traten auf Ebene der MedDRA SOC und PT in keinem Fall bei 5 % oder mehr der Patienten in einer Behandlungsgruppe auf, sodass von einer zusätzlichen Darstellung im Abschnitt 4.3.1.3.1.7 abgesehen wird.

### ***TEAE von besonderem Interesse***

Tabelle 4-59: Ergebnisse für *TEAE von besonderem Interesse* nach *System Organ Class (SOC)* und *Preferred Term (PT)* aus RCT mit dem zu bewertenden Arzneimittel (*Safety*-Population)

<b>UX023-CL303 Population</b>	<b>Patienten n (%)</b>	<b>OR<sup>a,b</sup> [95% KI]</b>	<b>p-Wert<sup>a</sup></b>	<b>RR<sup>b,c,e</sup> [95% KI]</b>	<b>ARD<sup>b,c,e</sup> [95% KI]</b>	<b>p-Wert<sup>d</sup></b>
<b>Baseline zu Woche 24</b>						
<b>Reaktionen an der Einstichstelle</b>						
Burosumab (68)	8 (11,8)	0,959 [0,333, 2,764]	0,9375	0,971 [0,387, 2,435]	0,004 [-0,106, 0,113]	1,0000
Placebo (66)	8 (12,1)					
<b>Hypersensitivität</b>						
Burosumab (68)	4 (5,9)	0,997 [0,232, 4,287]	0,9966	0,971 [0,253, 3,721]	0,002 [-0,078, 0,082]	1,0000
Placebo (66)	4 (6,1)					
<b>Hyperphosphatämie</b>						
Burosumab (68)	4 (5,9)	n.b.		8,739 [0,479, 159,198]	0,059 [0,003, 0,115]	0,1195
Placebo (66)	0 (0)					
<b>Restless-Legs-Syndrom</b>						
Burosumab (68)	8 (11,8)	1,381 [0,442, 4,314]	0,5765	1,294 [0,475, 3,528]	0,027 [-0,077, 0,130]	0,7791
Placebo (66)	6 (9,1)					
<p>N: Patientenzahl der entsprechend definierten Population; n: Anzahl der Patienten mit mindestens einem Ereignis in der entsprechenden Kategorie; %: Anteil von N; TEAE: <i>Treatment emergent adverse event</i> (unerwünschtes Ereignis nach Beginn der Behandlung); SOC: <i>System Organ Class</i> (Systemorganklasse); PT: <i>Preferred Term</i>, [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>; RR: <i>Risk Ratio</i> (relatives Risiko); ARD: <i>Absolute Risk Difference</i> (Absolute Risikodifferenz).</p> <p>Ein RR bzw. OR &lt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &gt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe und den tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) als unabhängige Variablen.</p> <p>b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p> <p>e: Treten keine Ereignisse auf, wird ein Korrekturwert von 0,5 zu den Zellen der Vierfeldertafel hinzugefügt.</p>						

Für alle *TEAE von besonderem Interesse* auf Ebene der MedDRA SOC und PT liegen keine statistisch signifikanten Unterschiede vor (Tabelle 4-59).

Zuvor durchgeführte Studien bei Erwachsenen mit XLH zeigten, dass aufgrund des Wirkmechanismus von Burosumab unter der Therapie eine Hyperphosphatämie auftreten kann. In der Studie UX023-CL303 traten bei insgesamt 5 Studienteilnehmern der Burosumab-Gruppe

während der 24-wöchigen doppelblinden Phase eine Hyperphosphatämie auf. Bei 4 der 5 Studienteilnehmer wurde die erhöhte Serumphosphat-Konzentration ( $> 4,5$  mg/dl) als unerwünschtes Ereignis der Kategorien „Hyperphosphatämie“ (PT) oder „Phosphor im Blut erhöht“ (PT) eingestuft. Die fünfte behandlungsbedürftige transiente Hyperphosphatämie wurde nicht als UE dokumentiert, da diese asymptomatisch verlief und sich die Serumphosphat-Konzentration innerhalb von 14 Tagen wieder normalisierte.

Es traten keine schwerwiegenden *TEAE von besonderem Interesse* oder  $\geq 3$  *TEAE von besonderem Interesse* auf.

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Für die Nutzenbewertung von Burosumab für Erwachsene mit XLH wurde keine Meta-Analyse durchgeführt, da im Anwendungsgebiet lediglich eine randomisierte kontrollierte Studie vorliegt, die den Einschlusskriterien in Abschnitt 4.2.2 entspricht.

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 wurde von Studienzentren in den USA und den europäischen Ländern Frankreich (drei Studienzentren), Italien (ein Studienzentrum), Irland (ein Studienzentrum) und Großbritannien (fünf Studienzentren) rekrutiert. Durch den Einschluss von europäischen Studienteilnehmern sowie der Tatsache, dass XLH eine genetisch bedingte Erkrankung ist, welche unabhängig von äußeren Einflüssen entsteht, liegen keine Anzeichen vor, die gegen eine Übertragbarkeit auf den deutschen Versorgungskontext schließen lassen [58]. Daher wird die doppelblinde, Placebo-kontrollierte und multizentrische Studie der Phase III (UX023-CL303) als übertragbar auf den deutschen Versorgungskontext angesehen. Zusätzlich wird eine weitere Studie, die BurGER-Studie, an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

#### 4.3.1.3.2 Subgruppenanalysen – RCT

Für die Darstellung der Ergebnisse aus Subgruppenanalysen gelten die gleichen Anforderungen wie für die Darstellung von Ergebnissen aus Gesamtpopulationen in Abschnitt 4.1.1.1.1.<sup>16</sup>

Darüber hinaus sind folgende Kriterien zu berücksichtigen:

- Subgruppenanalysen sind nur für die Merkmale (z.B. Alter) durchzuführen, bei denen die resultierenden Subgruppen jeweils mindestens 10 Patienten umfassen.
- Subgruppenanalysen sind für binäre Ereignisse je Merkmal nur dann durchzuführen, wenn in einer der Subgruppen mindestens 10 Ereignisse aufgetreten sind.
- Für Überlebenszeitanalysen müssen Kaplan-Meier-Kurven zu den einzelnen Subgruppen nur für Subgruppenanalysen mit statistisch signifikantem Interaktionsterm ( $p < 0,05$ ) dargestellt werden.
- Ergebnisse zu UE nach SOC und PT müssen nur dargestellt werden, wenn das jeweilige Ergebnis für die Gesamtpopulation statistisch signifikant ist. Zu a priori definierten Ereignissen (z.B. AESI, SMQs) sowie den UE-Gesamtraten (UE, schwere UE, SUE und Abbrüche wegen UE) müssen Subgruppenanalysen unabhängig vom Vorliegen statistischer Signifikanz in der Gesamtpopulation dargestellt werden.
- Bei Vorliegen mehrerer Studien und Durchführung von Metaanalysen zu diesen Studien gelten die zuvor genannten Kriterien für die jeweilige Metaanalyse, nicht für die Einzelstudien.
- Für Studien des pharmazeutischen Unternehmers sind entsprechende Analysen für alle benannten Effektmotifikatoren zu allen relevanten Endpunkten nach den zuvor genannten Kriterien vorzulegen und daher ggf. posthoc durchzuführen.
- Wird für die Nutzenbewertung nur die Teilpopulation einer Studie herangezogen (z.B. wegen Zulassungsbeschränkungen, aufgrund von durch den G-BA bestimmte Teilpopulationen), so gelten die genannten Kriterien für diese Teilpopulation, und die Subgruppenanalysen sind für die Teilpopulation und nicht für die Gesamtpopulation der Studie durchzuführen.
- Subgruppenanalysen, bei denen der Interaktionsterm nicht statistisch signifikant ist, können auch in einem separaten Anhang des vorliegenden Modul 4 dargestellt werden. Dabei kann die Ausgabe der Statistik-Software unverändert verwendet werden, sofern diese alle notwendigen Angaben enthält. Eine ausschließliche Darstellung in Modul 5 ist aber nicht ausreichend.

---

<sup>16</sup> unbesetzt

Beschreiben Sie die Ergebnisse von Subgruppenanalysen. Stellen Sie dabei zunächst tabellarisch dar, zu welchen der in Abschnitt 4.2.5.5 genannten Effektmodifikatoren Subgruppenanalysen zu den relevanten Endpunkten vorliegen, und ob diese a priori geplant und im Studienprotokoll festgelegt waren oder posthoc durchgeführt wurden.

Orientieren Sie sich an der beispielhaften Angabe in der ersten Tabellenzeile.

Die Subgruppenanalysen der doppelblinden Placebo-kontrollierten multizentrischen Studie UX023-CL303 umfassen sowohl präspezifizierte als auch *post-hoc* durchgeführte Analysen für alle patientenrelevante Endpunkte. Diese sind in Tabelle 4-60 und Tabelle 4-61 dargestellt.

Tabelle 4-60: Präspezifizierte Subgruppenanalysen

Endpunkt-Kategorie UX023-CL303	Morbidität						Sicherheit
	Serumphosphat	BPI	WOMAC	BFI	6MWT	PGI-I	
<b>Subgruppen präspezifiziert</b>							
<i>BPI Schlimmster Schmerz (&gt; 6,0 vs. ≤ 6,0)</i>	x	x	x	x	x	x	x
<i>BPI Durchschnittlicher Schmerz (&gt; 6,0 vs. ≤ 6,0)</i>	x	x	x	x	x	x	x
<i>Region (Nordamerika/EU vs. Japan vs. Südkorea)</i>	x	x	x	x	x	x	x
<i>Geschlecht (weiblich vs. männlich)</i>	x	x	x	x	x	x	x
<i>Ethnie (weiß vs. nicht weiß)</i>	x	x	x	x	x	x	x
<i>Alter (≤ 41 Jahre vs. &gt; 41 Jahre)</i>	x	x	x	x	x	x	x
<i>Verwendung von Schmerzmitteln zu Baseline (ja vs. nein)</i>	x	x	x	x	x	x	x
<i>Verwendung von Opioiden zu Baseline (ja vs. nein)</i>	x	x	x	x	x	x	x

Tabelle 4-61: *Post-hoc* durchgeführte Subgruppenanalysen

Endpunkt-Kategorie UX023-CL303	Morbidität						Sicherheit
	Serumphosphat	BPI	WOMAC	BFI	6MWT	PGI-I	
<b>Subgruppen <i>post-hoc</i> definiert</b>							
WOMAC Steifheit zu Baseline ( $\leq 62,5$ vs. $> 62,5$ )	x	x	x	x	x	x	x
WOMAC Physische Funktion zu Baseline ( $\leq 47,8$ vs. $> 47,8$ )	x	x	x	x	x	x	x
WOMAC Schmerzscore zu Baseline ( $\leq 50$ vs. $> 50$ )	x	x	x	x	x	x	x
WOMAC Gesamtscore zu Baseline ( $\leq 49,5$ vs. $> 49,5$ )	x	x	x	x	x	x	x
Frakturen zu Baseline (ja vs. nein)	x	x	x	x	x	x	x
6MWT zu Baseline ( $\leq 365$ m vs. $> 365$ m)	x	x	x	x	x	x	x

Stellen Sie anschließend in Tabelle 4-62 und Tabelle 4-63 die Ergebnisse der Interaktionsterme für alle Subgruppenanalysen je Endpunkt in tabellarischer Form dar, und zwar für jede einzelne Studie separat. Kennzeichnen Sie dabei statistisch signifikante ( $p < 0,05$ ) Interaktionsterme

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-62: p-Werte der Interaktion für präspezifizierte Subgruppenanalysen

Endpunkte		p-Werte der Interaktion							
Subgruppe	Endpunkt	BPI Durchschnittlicher Schmerz zu BL (> 6,0 vs. ≤ 6,0)	BPI Schlimmster Schmerz zu BL (> 6,0 vs. ≤ 6,0)	Region (Nordamerika/ EU vs. Japan vs. Südkorea)	Geschlecht (weiblich vs. männlich)	Ethnie (weiß vs. nicht weiß)	Alter (≤ 41 Jahre vs. > 41 Jahre)	Verwendung von Schmerzmitteln zu Baseline (ja vs. nein)	Verwendung von Opioiden zu Baseline (ja vs. nein)
		Datenschnitt	W24	W24	W24	W24	W24	W24	W24
<b>Morbidität</b>									
	Serumphosphat-Prozentsatz erreichter Normwert (≥ 2,5 mg/dl)	0,8520	0,9662	0,9275	0,8537	0,9126	0,5080	0,6432	0,9738
	Serumphosphat-Konzentration (mg/dl)	0,7964	0,6560	0,8092	0,3086	0,8659	0,0682	0,7858	0,9795
	BPI Schlimmster Schmerz	0,9945	0,6363	0,3919	0,4974	0,6245	0,6946	0,5203	0,7288
	BPI Schlimmster Schmerz 15 % Abnahme	0,2533	- <sup>a</sup>	0,4801	0,8009	0,9716	0,3184	0,5486	0,8651
	BPI Schlimmster Schmerz 30 % Abnahme	0,7967	- <sup>a</sup>	0,0734	0,6187	0,1337	0,6764	0,4288	0,2399
	BPI Schmerzinterferenz	0,1131	0,0829	0,4902	0,3078	0,8550	0,9201	0,7464	0,4449
	BPI Schmerzintensität	0,6390	0,9510	0,3623	0,5650	0,6153	0,4977	0,6018	0,4777
	WOMAC Steifheit	0,6635	0,5152	0,7852	0,1072	0,8047	0,7833	0,3772	0,6358
	WOMAC Steifheit Abnahme ≥ 15 nu	0,1882	0,4999	0,8559	0,6374	0,8849	0,8829	0,4431	0,9637
	WOMAC Physische Funktion	0,1837	0,2564	0,3521	0,3359	0,9236	0,2726	0,9098	0,7500
	WOMAC Physische Funktion Abnahme ≥ 15 nu	0,8779	0,2233	0,9586	0,5814	0,9488	0,4455	0,7819	0,8505
	BFI Score Schlimmste Fatigue	0,8583	0,1802	0,7375	0,6445	0,7066	0,9851	0,1845	0,5520
	BFI Globaler Fatigue Score	0,4794	0,7120	0,6178	0,2534	0,7427	0,2413	0,6288	0,3403

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Endpunkte		p-Werte der Interaktion						
Subgruppe	BPI Durchschnittlicher Schmerz zu BL (> 6,0 vs. ≤ 6,0)	BPI Schlimmster Schmerz zu BL (> 6,0 vs. ≤ 6,0)	Region (Nordamerika/ EU vs. Japan vs. Südkorea)	Geschlecht (weiblich vs. männlich)	Ethnie (weiß vs. nicht weiß)	Alter (≤ 41 Jahre vs. > 41 Jahre)	Verwendung von Schmerzmitteln zu Baseline (ja vs. nein)	Verwendung von Opioiden zu Baseline (ja vs. nein)
Endpunkt	W24	W24	W24	W24	W24	W24	W24	W24
Datenschnitt	W24	W24	W24	W24	W24	W24	W24	W24
6MWT (tatsächliche Gesamtdistanz, korrigierte Analyse)	0,5123	<b>0,0075</b>	0,1703	0,6149	0,0767	0,0960	0,8973	0,8990
6MWT (tatsächliche Gesamtdistanz, Sensitivitätsanalyse)	0,5371	<b>0,0085</b>	0,1641	0,5903	0,0725	0,0849	0,8713	0,9250
6MWT (Prozentsatz, korrigierte Analyse)	0,5184	<b>0,0159</b>	0,2051	0,5296	0,0829	0,0681	0,9674	0,9048
6MWT (Prozentsatz, Sensitivitätsanalyse)	0,5422	<b>0,0175</b>	0,1979	0,5048	0,0786	0,0590	0,9403	0,9280
PGI-I	0,8436	0,4770	0,2609	0,6446	0,3397	0,7077	0,2823	0,3146
<b>Sicherheit</b>								
<b>Unerwünschte Ereignisse nach Beginn der Behandlung (TEAE)</b>								
TEAE	0,9180	0,4859	0,9743	0,3698	0,9720	0,4829	0,2828	0,9753
Schwerwiegende TEAE	_b	_b	_b	_b	_b	_b	_b	_b
TEAE Grad 3 – 4	0,1867	0,8353	0,9998	0,5135	0,9038	0,5046	0,7154	0,9136
<b>TEAE von besonderem Interesse</b>								
Reaktionen an der Einstichstelle	0,1702	0,7441	0,6018	0,9954	0,9997	0,1136	0,4091	0,1528
Restless-Legs-Syndrom	0,9443	0,7974	0,7809	0,2262	0,8786	0,2842	0,3872	0,4771
<b>TEAE nach MedDRA SOC und PT (Inzidenz ≥ 10 % in mindestens einer Behandlungsgruppe)</b>								
PT Arthralgie	0,4847	0,5967	0,6454	0,3708	0,1055	0,5161	0,5472	0,4992

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Endpunkte		p-Werte der Interaktion							
Subgruppe	Endpunkt	BPI Durchschnittlicher Schmerz zu BL (> 6,0 vs. ≤ 6,0)	BPI Schlimmster Schmerz zu BL (> 6,0 vs. ≤ 6,0)	Region (Nordamerika/ EU vs. Japan vs. Südkorea)	Geschlecht (weiblich vs. männlich)	Ethnie (weiß vs. nicht weiß)	Alter (≤ 41 Jahre vs. > 41 Jahre)	Verwendung von Schmerzmitteln zu Baseline (ja vs. nein)	Verwendung von Opioiden zu Baseline (ja vs. nein)
	Datenschnitt	W24	W24	W24	W24	W24	W24	W24	W24

BPI: Brief Pain Inventory; WOMAC: Western Ontario and McMaster Universities Osteoarthritis; BFI: Brief Fatigue Inventory; 6MWT: 6-Minuten-Gehdistanz Test; PGI-I: Patient Global Impression of Improvement; TEAE: unerwünschtes Ereignis nach Beginn der Behandlung (Treatment emergent adverse event); SOC: System Organ Class; PT: Preferred Term

a: nicht analysiert  
 b: erfüllt nicht die Anforderungen an die Subgruppenanalysen in Abschnitt 4.3.1.3.2

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-63: p-Werte der Interaktion für *post-hoc* durchgeführte Subgruppenanalysen

Endpunkte		p-Werte der Interaktion					
Subgruppe	Endpunkt	WOMAC Steifheit zu Baseline ( $\leq 62,5$ vs. $> 62,5$ )	WOMAC Physische Funktion zu Baseline ( $\leq 47,8$ vs. $> 47,8$ )	WOMAC Schmerzscore zu Baseline ( $\leq 50$ vs. $> 50$ )	WOMAC Gesamtscore zu Baseline ( $\leq 49,5$ vs. $> 49,5$ )	Aktive Frakturen und Pseudofrakturen zu Baseline (ja vs. nein)	6MWT zu Baseline ( $\leq 365m$ vs. $> 365m$ )
		Datenschnitt	W24	W24	W24	W24	W24
<b>Morbidität</b>							
	Serumphosphat-Prozentsatz erreichter Normwert ( $\geq 2,5$ mg/dl)	0,9303	0,4678	0,1514	0,4678	0,9461	0,9582
	Serumphosphat-Konzentration (mg/dl)	0,6584	0,3966	0,9696	0,4907	0,8582	0,6626
	BPI Schlimmster Schmerz	0,9393	0,9643	0,1317	0,9517	0,4938	0,2973
	BPI Schlimmster Schmerz 15 % Abnahme	0,8251	0,8971	0,1619	0,6382	0,7754	0,2516
	BPI Schlimmster Schmerz 30 % Abnahme	0,6127	0,1659	0,5985	0,1659	0,8532	0,1120
	BPI Schmerzinterferenz	0,1528	0,3859	0,8916	0,1754	0,6372	0,3277
	BPI Schmerzintensität	0,6138	0,9958	0,3423	0,8076	0,4208	0,3813
	WOMAC Steifheit	0,8437	0,1857	0,1013	0,6912	0,3038	0,2382
	WOMAC Steifheit Abnahme $\geq 15$ nu	0,3975	<b>0,0474</b>	0,1958	<b>0,0231</b>	0,5939	<b>0,0278</b>
	WOMAC Physische Funktion	0,0796	0, 4164	0,4464	0,2121	0,2644	0,7468
	WOMAC Physische Funktion Abnahme $\geq 15$ nu	0,1165	0,6790	0,6655	0,3159	0,4965	0,3486
	BFI Score Schlimmste Fatigue	0,9474	0,5041	0,4262	0,4807	0,8562	0,7887
	BFI Globaler Fatigue Score	0,1257	0,3407	0,7989	0,1226	0,6580	0,5836

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Endpunkte		p-Werte der Interaktion				
Subgruppe	WOMAC Steifheit zu Baseline (≤ 62,5 vs. > 62,5)	WOMAC Physische Funktion zu Baseline (≤ 47,8 vs. > 47,8)	WOMAC Schmerzscore zu Baseline (≤ 50 vs. > 50)	WOMAC Gesamtscore zu Baseline (≤ 49,5 vs. > 49,5)	Aktive Frakturen und Pseudofrakturen zu Baseline (ja vs. nein)	6MWT zu Baseline (≤ 365m vs. > 365m)
Endpunkt						
<b>Datenschnitt</b>	W24	W24	W24	W24	W24	W24
6MWT (tatsächliche Gesamtdistanz, korrigierte Analyse)	0,7384	0,4072	0,3985	0,3351	0,8544	0,6641
6MWT (tatsächliche Gesamtdistanz, Sensitivitätsanalyse)	0,6777	0,4489	0,4300	0,3127	0,8003	0,6134
6MWT (Prozentsatz, korrigierte Analyse)	0,7578	0,4979	0,3133	0,4156	0,9436	0,7736
6MWT (Prozentsatz, Sensitivitätsanalyse)	0,6970	0,5411	0,3411	0,3877	0,8925	0,7200
PGI-I	0,4720	0,3708	<b>0,0100</b>	0,6005	0,6053	0,6871
<b>Sicherheit</b>						
<b>Unerwünschte Ereignisse nach Beginn der Behandlung (TEAE)</b>						
TEAE	0,9768	0,9133	0,9612	0,9990	0,9699	0,1538
Schwerwiegende TEAE	_a	_a	_a	_a	_a	_a
TEAE Grad 3 und 4	0,9812	0,3685	0,7090	0,3275	0,8962	0,8903
<b>TEAE von besonderem Interesse</b>						
Reaktionen an der Einstichstelle	0,8123	0,5496	0,4537	0,5496	0,2109	0,1206
Restless-Legs-Syndrom	0,3593	0,4568	0,5034	0,4568	0,5190	0,4771
<b>TEAE nach MedDRA SOC und PT (Inzidenz ≥ 10 % in mindestens einer Behandlungsgruppe)</b>						
PT Arthralgie	0,7510	0,3459	0,2416	0,3459	0,2562	0,9175
BPI: Brief Pain Inventory; WOMAC: Western Ontario and McMaster Universities Osteoarthritis; BFI: Brief Fatigue Inventory; 6MWT: 6-Minuten-Gehdistanz Test; PGI-I: Patient						

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Endpunkte		p-Werte der Interaktion				
Subgruppe	WOMAC Steifheit zu Baseline ( $\leq 62,5$ vs. $> 62,5$ )	WOMAC Physische Funktion zu Baseline ( $\leq 47,8$ vs. $> 47,8$ )	WOMAC Schmerzscore zu Baseline ( $\leq 50$ vs. $> 50$ )	WOMAC Gesamtscore zu Baseline ( $\leq 49,5$ vs. $> 49,5$ )	Aktive Frakturen und Pseudofrakturen zu Baseline (ja vs. nein)	6MWT zu Baseline ( $\leq 365m$ vs. $> 365m$ )
Endpunkt						
Datenschnitt	W24	W24	W24	W24	W24	W24
<i>Global Impression of Improvement</i> ; TEAE: unerwünschtes Ereignis nach Beginn der Behandlung ( <i>Treatment emergent adverse event</i> ); SOC: <i>System Organ Class</i> ; PT: <i>Preferred Term</i> a: erfüllt nicht die Anforderungen an die Subgruppenanalysen in Abschnitt 4.3.1.3.2						

*Stellen Sie schließlich alle Subgruppenergebnisse dar.*

*Sofern eine Effektmodifikation für mehr als ein Subgruppenmerkmal vorliegt, kann eine Untersuchung auf eine Wechselwirkung höherer Ordnung sinnvoll sein. Dies gilt insbesondere dann, wenn diese Effektmodifikation konsistent über mehrere Endpunkte besteht. Zur Interpretation der Ergebnisse sollte dann für diese Endpunkte zusätzlich eine Subgruppenanalyse durchgeführt werden, die die Merkmale mit Effektmodifikation kombiniert. Beispiel: Für die Endpunkte Mortalität, gesundheitsbezogene Lebensqualität und schwere unerwünschte Ereignisse liegt sowohl für das Merkmal Geschlecht (mit den Ausprägungen „weiblich“ und „männlich“) als auch für das Merkmal Schweregrad (mit den Ausprägungen „niedrig“ und „hoch“) eine Effektmodifikation vor. Die zusätzliche Subgruppenanalyse erfolgt dann für die 3 genannten Endpunkte für das kombinierte Merkmal Geschlecht/Schweregrad mit den 4 Ausprägungen weiblich/niedrig, weiblich/hoch, männlich/niedrig und männlich/hoch.*

*Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Ergebnisse mithilfe einer Meta-Analyse quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analyse (als Forest-Plot) dar.*

*Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Meta-Analyse durchführen bzw. wenn Sie nicht alle Studien in die Meta-Analyse einschließen.*

Die Analyse von Effektmodifikatoren für Subgruppenmerkmale in der doppelblinden Placebo-kontrollierten multizentrischen Studie UX023-CL303 fand wie in Abschnitt 4.2.5.5 beschrieben anhand eines mehrstufigen Prozesses statt und beruht auf Interaktions- bzw. Homogenitätstests. Hierbei ist zu beachten, dass aufgrund des multiplen Testens wie hier bei der Durchführung des Interaktionstests eine gewisse Anzahl der Tests bereits aufgrund der angenommenen Irrtumswahrscheinlichkeit von 5 % zu einem falsch positiven Ergebnis führt.

Die TEAE von besonderem Interesse Hypersensitivität und Hyperphosphatämie erfüllten nicht die in Abschnitt 4.3.1.3.2 erforderlichen Kriterien zur Darstellung von Subgruppenanalysen, sodass von einer Berechnung der Interaktions-p-Werte und Subgruppenanalysen abgesehen wird.

#### **4.3.1.3.2.1 Ergebnisse für präspezifizierte Subgruppenanalysen**

Im Folgenden werden alle Subgruppenanalysen, für die ein positiver Interaktionstest ( $p < 0,05$ ) vorliegt, detailliert dargestellt. Entsprechend des in Abschnitt 4.2.5.5 beschriebenen strukturierten, mehrstufigen Vorgehen wird im nächsten Schritt anhand der Lage des Effektschätzers und des KI geprüft, ob eine abweichende Interpretation des Zusatznutzens und damit eine bewertungsrelevante Effektmodifikation grundsätzlich vorliegen könnte. Ist dies der Fall, wird abschließend anhand der in Abschnitt 4.2.5.5 beschriebenen Detailkriterien untersucht, ob es sich um eine tatsächliche Effektmodifikation oder etwa um ein falsch positives Ergebnis infolge des multiplen Testens und der geringen Patientenzahl in den Behandlungsgruppen handelt.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

### 4.3.1.3.2.1.1 Gehfähigkeit

Tabelle 4-64: Subgruppenanalysen – Ergebnisse für *Gehfähigkeit* aus RCT mit dem zu bewertenden Arzneimittel

Intervention	<i>6MWT (6-Minuten-Gehdistanz Test)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MWD (SE) LsMean <sup>a</sup> (SE)	n MWD (SE) LsMean <sup>a</sup> (SE)	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
<i>Korrigierte Analyse Gesamtrate – Messung in Meter<sup>b</sup></i>								
<i>Subgruppe BPI Schlimmster Schmerz zu BL ≤ 6,0</i>								
Placebo (N=27)	27 417,26 (100,269)	27 417,26 (100,269)	27 406,41 (90,967)	27 417,26 (100,269)	27 400,44 (97,746)	27 -10,85 (5,177) -25,58 (6,204)	27 -16,81 (7,009) -31,54 (7,638)	42,95 [18,37; 67,52] 0,0006 H'g = 1,110 [0,448; 1,771]
Burosumab (N=17)	16 383,63 (112,350)	16 383,63 (112,350)	16 416,44 (100,140)	16 383,63 (112,350)	16 415,31 (100,213)	16 32,81 (8,261) 12,53 (7,993)	16 31,69 (12,494) 11,40 (11,095)	
<i>Subgruppe BPI Schlimmster Schmerz zu BL &gt; 6,0</i>								
Placebo (N=39)	38 332,00 (91,271)	37 333,43 (92,096)	37 345,03 (98,685)	37 330,81 (92,232)	37 346,81 (102,749)	37 11,59 (7,513) 4,04 (14,351)	37 16,00 (6,929) 7,16 (13,643)	6,28 [-13,65; 26,21] 0,5371 -
Burosumab (N=51)	51 348,35 (108,291)	48 346,63 (110,984)	48 359,81 (104,920)	49 351,63 (108,734)	49 370,39 (109,729)	48 13,19 (6,267) 7,87 (11,527)	49 18,76 (6,965) 13,44 (11,608)	

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention	6MWT (6-Minuten-Gehdistanz Test)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup>
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	[95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
<b>Korrigierte Analyse Gesamrate – Prozentsatz<sup>b</sup></b>								
<b>Subgruppe BPI Schlimmster Schmerz zu BL ≤ 6,0</b>								
Placebo (N=27)	27 59,75 (13,242)	27 59,75 (13,242)	27 58,38 (12,463)	27 59,75 (13,242)	27 57,43 (12,734)	27 -1,37 (0,760) -3,79 (0,897)	27 -2,32 (0,986) -4,74 (1,091)	6,27 [2,71; 9,83] 0,0006 H'g = 1,123 [0,460; 1,785]
Burosumab (N=17)	16 55,30 (17,584)	16 55,30 (17,584)	16 59,97 (15,861)	16 55,30 (17,584)	16 59,96 (16,167)	16 4,67 (1,188) 1,55 (1,095)	16 4,66 (1,796) 1,53 (1,588)	
<b>Subgruppe BPI Schlimmster Schmerz zu BL &gt; 6,0</b>								
Placebo (N=39)	38 46,92 (13,806)	37 47,10 (13,947)	37 48,57 (13,850)	37 46,63 (13,878)	37 48,77 (14,482)	37 1,47 (1,047) 0,10 (1,876)	37 2,15 (0,947) 0,58 (1,778)	1,43 [-1,40; 4,26] 0,3235
Burosumab (N=51)	51 50,22 (15,221)	48 49,98 (15,580)	48 52,04 (15,034)	49 50,75 (15,208)	49 53,63 (15,605)	48 2,06 (0,913) 1,17 (1,504)	49 2,89 (1,027) 2,00 (1,552)	

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention	6MWT (6-Minuten-Gehdistanz Test)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MWD (SE) LsMean <sup>a</sup> (SE)	n MWD (SE) LsMean <sup>a</sup> (SE)	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
<b>Sensitivitätsanalyse Gesamtrate – Messung in Meter<sup>c</sup></b>								
<b>Subgruppe BPI Schlimmster Schmerz zu BL ≤ 6,0</b>								
Placebo (N=27)	27 417,26 (100,269)	27 417,26 (100,269)	27 406,41 (90,967)	27 417,26 (100,269)	27 400,44 (97,746)	27 -10,85 (5,177) -25,58 (6,204)	27 -16,81 (7,009) -31,54 (7,638)	42,95 [18,37; 67,52] 0,0006 H'g = 1,110 [0,448; 1,771]
Burosumab (N=17)	16 383,63 112,350)	16 383,63 (112,350)	16 416,44 (101,140)	16 383,63 (112,350)	16 415,31 (100,213)	16 32,81(8,261) 12,53 (7,993)	16 31,69 (12,494) 11,40 (11,095)	
<b>Subgruppe BPI Schlimmster Schmerz zu BL &gt; 6,0</b>								
Placebo (N=39)	38 332,00 (91,271)	37 333,43 (92,096)	37 345,03 (98,685)	37 330,81 (92,232)	37 346,81 (102,749)	37 11,59 (7,513) 3,93 (14,371)	37 16,00 (6,929) 7,00 (13,671)	7,91 [-11,91; 27,73] 0,4343
Burosumab (N=51)	50 346,12 (108,198)	47 344,21 (110,905)	47 357,00 (104,210)	48 349,38 (108,717)	48 369,90 (110,835)	47 12,79 (6,388) 7,15 (11,543)	48 20,52 (6,880) 14,91 (11,578)	

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention	6MWT (6-Minuten-Gehdistanz Test)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MW (SD)	n MWD (SE) LsMean <sup>a</sup> (SE)	n MWD (SE) LsMean <sup>a</sup> (SE)	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
<b>Sensitivitätsanalyse Gesamtrate – Prozentsatz<sup>c</sup></b>								
<b>Subgruppe BPI Schlimmster Schmerz zu BL ≤ 6,0</b>								
Placebo (N=27)	27 59,75 (13,242)	27 59,75 (13,242)	27 58,38 (12,463)	27 59,75 (13,242)	27 57,43 (12,734)	27 -1,37 (0,760) -3,79 (0,897)	27 -2,32 (0,986) -4,74 (1,091)	6,27 [2,71; 9,83] 0,0006 H'g = 1,123 [0,460; 1,785]
Burosumab (N=17)	16 55,30 (17,584)	16 55,30 (17,584)	16 59,97 (15,861)	16 55,30 (17,584)	16 59,96 (16,167)	16 4,67 (1,188) 1,55 (1,095)	16 4,66 (1,796) 1,53 (1,588)	
<b>Subgruppe BPI Schlimmster Schmerz zu BL &gt; 6,0</b>								
Placebo (N=39)	38 46,92 (13,806)	37 47,10 (13,947)	37 48,57 (13,850)	37 46,63 (13,878)	37 48,77 (14,482)	37 1,47 (1,047) 0,10 (1,877)	37 2,15 (0,947) 0,57 (1,781)	1,66 [-1,16; 4,47] 0,2490 -
Burosumab (N=51)	50 49,85 (15,144)	47 49,58 (15,501)	47 51,58 (14,856)	48 50,37 (15,141)	48 53,52 (15,751)	47 2,00 (0,930) 1,07 (1,505)	48 3,15 (1,013) 2,23 (1,546)	
<p>MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; %: Prozentanteil von N; [95% KI]: 95%-iges Konfidenzintervall; SE: <i>standard error</i> (Standardfehler); H'g: Hedges'g (wurde <i>post-hoc</i> berechnet), BL: <i>Baseline</i></p> <p>a: Berechnung mit <i>generalized estimation equation</i> (GEE)-Modell zu Woche 96.</p> <p>b: Der fehlerhaft kalkulierte Baseline-Wert eines Studienteilnehmers wurde korrigiert und in die Analyse einbezogen.</p> <p>c: Analyse beinhaltet nicht den Studienteilnehmer mit dem fehlerhaft kalkulierten Baseline-Wert.</p> <p>GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts <i>BPI Durchschnittlichen Schmerz</i>, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit der Änderung von Baseline des <i>6MWT</i> als abhängige Variable und Region, Visite, Behandlungsarm, <i>BPI</i></p>								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention	<i>6MWT (6-Minuten-Gehdistanz Test)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<p><i>Durchschnittlicher Schmerz</i> (Stratifikationsfaktor) und <i>Visite/Behandlungsarm</i> Interaktion als feste Effekte. Der Baseline-Wert des <i>6MWT</i> ist als zusätzliche Kovariable eingefügt. Als Kovarianzstruktur wird eine <i>Compound Symmetry</i> Matrix verwendet.</p>								

Für den Endpunkt *Gehfähigkeit* (6MWT – Gesamtrate Messung in Meter, korrigierte Analyse) lässt sich eine potenzielle Effektmodifikation in der Subgruppe *BPI Schlimmster Schmerz* zu Baseline ( $\leq 6,0$  vs.  $> 6,0$ ) beobachten. Bei einem p-Wert im Interaktionstest von  $p = 0,0075$  ergibt sich für die Ausprägung *BPI Schlimmster Schmerz* zu Baseline  $\leq 6,0$  eine statistisch signifikante Verbesserung der Gehfähigkeit (W24: Differenz $_{\leq 6,0}$ : 42,95 m;  $p = 0,0006$ ;  $H^* = 1,110 [0,448; 1,771]$ ), die insgesamt stärker ausgebildet ist als bei der Ausprägung *BPI Schlimmster Schmerz* zu Baseline  $> 6,0$  (W24: Differenz $_{> 6,0}$ : 6,28 m;  $p = 0,5371$ ) (Tabelle 4-64). Zu beachten ist jedoch, dass, obwohl die Effektschätzer gleichgerichtet sind, lediglich bei der Ausprägung *BPI Schlimmster Schmerz* zu Baseline  $\leq 6,0$  eine statistische Signifikanz vorliegt.

Des Weiteren ist für die gleiche Subgruppe eine potenzielle Effektmodifikation bei dem Endpunkt *Gehfähigkeit* (Gesamtrate – Prozentsatz, korrigierte Analyse) vorhanden. Die Effektschätzer sind ebenfalls gleichgerichtet und unterscheiden sich, wie bereits oben beschrieben, anhand ihrer statistischen Signifikanz (W24: Differenz $_{\leq 6,0}$ : 6,27 m;  $p = 0,0006$ ;  $H^*g = 1,123 [0,460; 1,785]$ ; Differenz $_{> 6,0}$ : 1,43 m;  $p = 0,3235$ ) (Tabelle 4-64).

Die Subgruppenanalyse der Sensitivitätsanalyse des 6MWT, die den Studienteilnehmer mit dem fehlerhaft kalkulierten Baseline-Wert ausschloss, ergab im Vergleich zur Primäranalyse kein abweichendes Ergebnis (Tabelle 4-64).

Da die Effektschätzer in beiden Subgruppen gleichgerichtet sind, wird nicht von einer relevanten Effektmodifikation ausgegangen. Zur Ableitung des Zusatznutzens werden die Ergebnisse der Gesamtpopulation herangezogen.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

### 4.3.1.3.2.2 Ergebnisse für *post-hoc* durchgeführte Subgruppenanalysen

#### 4.3.1.3.2.2.1 Erfassung von Therapieeffekten

Tabelle 4-65: Subgruppenanalysen – Ergebnisse für *Erfassung von Therapieeffekten* aus RCT mit dem zu bewertenden Arzneimittel

Intervention (N)	PGI-I (Patient Global Impression of Improvement)							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	n	n	n	n	n	n	n	LsMean <sup>a</sup> [95% KI] p-Wert falls p < 0,05: Hedges'g [95% KI]
	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MW (SD)	MWD (SE) LsMean <sup>a</sup> (SE)	MWD (SE) LsMean <sup>a</sup> (SE)	
<b>Subgruppe WOMAC Schmerzscore zu BL ≤ 50</b>								
Placebo (N=42)	42 2,71 (0,805)	42 2,71 (0,805)	42 3,76 (1,031)	41 2,73 (0,807)	41 3,98 (1,332)	_ <sup>b</sup> _ <sup>b</sup> 3,94 (0,168)	_ <sup>b</sup> _ <sup>b</sup> 4,14 (0,220)	-0,67 [-1,21; -0,14] 0,0131 H'g = -0,556 [-1,027; -0,085]
Burosumab (N=32)	32 2,94 (0,619)	32 2,94 (0,619)	32 3,19 (0,998)	32 2,94 (0,619)	32 3,28 (1,054)	_ <sup>b</sup> _ <sup>b</sup> 3,38 (0,210)	_ <sup>b</sup> _ <sup>b</sup> 3,47 (0,222)	
<b>Subgruppe WOMAC Schmerzscore zu BL &gt; 50</b>								
Placebo (N=24)	24 3,50 (0,590)	24 3,50 (0,590)	24 3,33 (1,239)	23 3,48 (0,593)	23 3,61 (1,500)	_ <sup>b</sup> _ <sup>b</sup> 3,34 (0,243)	_ <sup>b</sup> _ <sup>b</sup> 3,60 (0,294)	0,22 [-0,51; 0,96] 0,5507 -
Burosumab (N=25)	35 3,40 (0,553)	34 3,38 (0,551)	34 3,79 (1,250)	34 3,38 (0,551)	34 3,79 (1,366)	_ <sup>b</sup> _ <sup>b</sup> 3,78 (0,210)	_ <sup>b</sup> _ <sup>b</sup> 3,82 (0,232)	
MWD: Mittelwertdifferenz; N: Anzahl der Patienten in der Studie; n: Anzahl der Patienten, die analysiert worden sind; [95% KI]: 95%-iges Konfidenzintervall; SE: <i>standard error</i> (Standardfehler); H'g: Hedges'g (wurde <i>post-hoc</i> berechnet), BL: <i>Baseline</i>								

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Intervention (N)	<i>PGI-I (Patient Global Impression of Improvement)</i>							
	Baseline	Baseline Woche 12	Woche 12	Baseline Woche 24	Woche 24	Änderung von Baseline zu Woche 12	Änderung von Baseline zu Woche 24	Behandlungsdifferenz von Baseline zu Woche 24
	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>LsMean<sup>a</sup> [95% KI]</b>
	<b>MW (SD)</b>	<b>MW (SD)</b>	<b>MW (SD)</b>	<b>MW (SD)</b>	<b>MW (SD)</b>	<b>MWD (SE) LsMean<sup>a</sup> (SE)</b>	<b>MWD (SE) LsMean<sup>a</sup> (SE)</b>	<b>p-Wert falls p &lt; 0,05: Hedges'g [95% KI]</b>

a: Berechnung mit *generalized estimation equation* (GEE)-Modell zu Woche 96.  
 b: Nicht bestimmt.  
 GEE: Die tatsächliche Stratifizierung innerhalb der Randomisierung basiert auf dem Mittelwert des Endpunkts *BPI Durchschnittlichen Schmerz*, berechnet sieben Tage vor der Baseline-Visite. Die Effektschätzer basieren auf einem GEE-Modell mit *PGI* als abhängige Variable und Region, Visite, Behandlungsarm, *BPI Durchschnittlicher Schmerz* (Stratifikationsfaktor) und Visite/Behandlungsarm Interaktion als feste Effekte. Als Kovarianzstruktur wird eine *Compound Symmetry* Matrix verwendet.

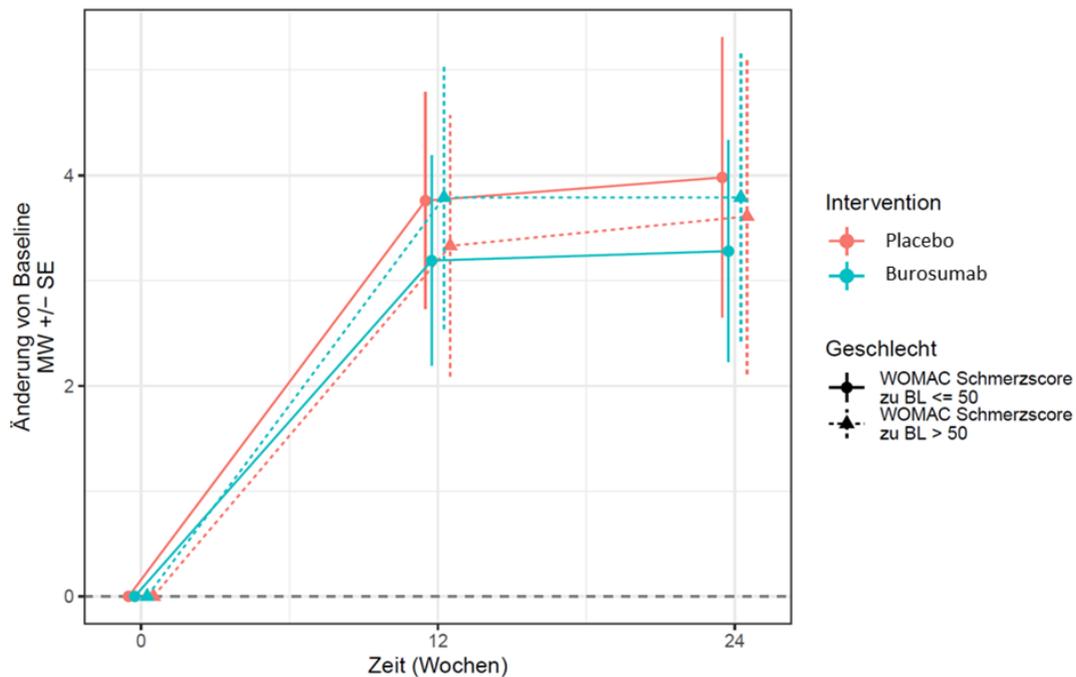


Abbildung 8: Subgruppenanalyse PGI-C

Für den Endpunkt *Erfassung von Therapieeffekten* kann eine potenzielle Effektmodifikation in der Subgruppe *WOMAC Schmerzscore zu Baseline ( $\leq 50,0$  vs.  $> 50,0$ )* beobachtet werden. Bei einem p-Wert im Interaktionstest von  $p = 0,0100$  (Woche 24) profitieren Studienteilnehmer mit der Ausprägung *WOMAC Schmerzscore zu Baseline  $\leq 50,0$*  im Vergleich zur Ausprägung *WOMAC Schmerzscore zu Baseline  $> 50,0$*  mehr von einer Behandlung mit Burosumab. Während in der Ausprägung *WOMAC Schmerzscore zu Baseline  $\leq 50,0$*  ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen vorliegt (Woche 24: Differenz<sub>≤47,8</sub>: -0,67;  $p = 0,0131$ ), ist der Unterschied zwischen den Behandlungsgruppen für die Ausprägung *WOMAC Schmerzscore zu Baseline  $> 50,0$*  (W<sub>24</sub>: Differenz<sub>≤47,8</sub>: 0,22;  $p = 0,5507$ ) nicht signifikant (Tabelle 4-65). Aufgrund der unterschiedlichen Signifikanz der Effektschätzer erfolgt die Ableitung des Zusatznutzens ausschließlich aus den Ergebnissen für die Gesamtpopulation.

**4.3.1.3.2.2 Steifheit**Tabelle 4-66: Subgruppenanalysen – Ergebnisse für *WOMAC Steifheit* aus RCT mit dem zu bewertenden Arzneimittel (Subgruppe *WOMAC Physische Funktion*)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p-Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p-Wert <sup>d</sup>
<b>WOMAC Physische Funktion ≤ 47,8</b>						
Burosumab (30)	7 (23,33)	1,103 [0,341, 3,571]	0,8678	1,079 [0,442; 2,635]	0,017 [-0,184; 0,218]	1,000
Placebo (37)	8 (21,62)					
<b>WOMAC Physische Funktion &gt; 47,8</b>						
Burosumab (38)	14 (36,84)	7,875 [1,574, 39,409]	0,0128	5,342 [1,317; 21,675]	0,299 [0,120; 0,478]	0,0079
Placebo (29)	2 (6,90)					
<p>N: Anzahl der Patienten in der Studie; n: Anzahl der Responder; %: Anteil von N; [95% KI]: 95%-iges Konfidenzintervall; OR: <i>Odds Ratio</i>; RR: <i>Risk Ratio</i> (relatives Risiko); ARD: <i>Adjusted Risk Difference</i> (Absolute Risikodifferenz), nu: <i>normalised units</i></p> <p>Ein RR bzw. OR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &lt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe als unabhängige Variable.</p> <p>b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>						

Für den Endpunkt *WOMAC Steifheit* kann eine potenzielle Effektmodifikation in der Subgruppe *WOMAC Physische Funktion zu Baseline* ( $\leq 47,8$  vs.  $> 47,8$ ) beobachtet werden. Bei einem p-Wert im Interaktionstest von  $p = 0,0474$  (Woche 24) profitieren Studienteilnehmer mit der Ausprägung *WOMAC Physische Funktion zu Baseline*  $> 47,8$  im Vergleich zur Ausprägung *WOMAC Physische Funktion zu Baseline*  $\leq 47,8$  mehr von einer Behandlung mit Burosumab. Während in der Ausprägung *WOMAC Physische Funktion zu Baseline*  $> 47,8$  ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen vorliegt (Woche 24: RR<sub>>47,8</sub> [95% KI]: 5,342; [1,317; 21,675];  $p = 0,0079$ ), ist der Unterschied zwischen den Behandlungsgruppen für die Ausprägung *WOMAC Physische Funktion zu Baseline*  $\leq 47,8$  (W<sub>24</sub>: RR<sub>≤47,8</sub> [95% KI]: 1,079; [0,442; 2,635];  $p = 1,000$ ) nicht signifikant (Tabelle 4-66). Aufgrund dessen erfolgt die Ableitung des Zusatznutzens ausschließlich aus den Ergebnissen für die Gesamtpopulation.

Tabelle 4-67: Subgruppenanalysen – Ergebnisse für *WOMAC Steifheit* aus RCT mit dem zu bewertenden Arzneimittel (Subgruppe *WOMAC Gesamtscore*)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p-Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p-Wert <sup>d</sup>
<b>WOMAC Gesamtscore ≤ 49,5</b>						
Burosumab (30)	6 (20,00)	0,906 [0,270; 3,043]	0,8716	0,925 [0,360; 2,374]	-0,016 [-0,211; 0,179]	1,000
Placebo (37)	8 (21,62)					
<b>WOMAC Gesamtscore &gt; 49,5</b>						
Burosumab (38)	15 (39,47)	8,804 [1,766; 43,899]	0,0087	5,724 [1,420; 23,072]	0,326 [0,145; 0,506]	0,0038
Placebo (29)	2 (6,90)					
<p>N: Anzahl der Patienten in der Studie; n: Anzahl der Responder; %: Anteil von N; [95% KI]: 95%-iges Konfidenzintervall;  OR: <i>Odds Ratio</i>, RR: <i>Risk Ratio</i> (relatives Risiko); ARD: <i>Adjusted Risk Difference</i> (Absolute Risikodifferenz), nu:  <i>normalised units</i></p> <p>Ein RR bzw. OR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR &lt; 1 einen Vorteil für Placebo.</p> <p>a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe als unabhängige Variable.</p> <p>b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte <i>post-hoc</i>.</p> <p>c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.</p> <p>d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.</p>						

Für den Endpunkt *WOMAC Steifheit* kann eine potenzielle Effektmodifikation in der Subgruppe *WOMAC Gesamtscore zu Baseline* ( $\leq 49,5$  vs.  $> 49,5$ ) beobachtet werden. Bei einem p-Wert im Interaktionstest von  $p = 0,0231$  (Woche 24) profitieren Studienteilnehmer mit der Ausprägung *WOMAC Gesamtscore zu Baseline*  $> 49,5$  im Vergleich zur Ausprägung *WOMAC Gesamtscore zu Baseline*  $\leq 49,5$  mehr von einer Behandlung mit Burosumab. Während in der Ausprägung *WOMAC Gesamtscore zu Baseline*  $> 49,5$  ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen vorliegt (Woche 24: RR<sub>>49,5</sub> [95% KI]: 5,724; [1,420; 23,072];  $p = 0,0038$ ), ist der Unterschied zwischen den Behandlungsgruppen für die Ausprägung *WOMAC Gesamtscore zu Baseline*  $\leq 49,5$  (W<sub>24</sub>: RR<sub>≤49,5</sub> [95% KI]: 0,925; [0,360; 2,374];  $p = 1,000$ ) nicht signifikant (Tabelle 4-67). Aufgrund dessen erfolgt die Ableitung des Zusatznutzens ausschließlich aus den Ergebnissen für die Gesamtpopulation.

Tabelle 4-68: Subgruppenanalysen – Ergebnisse für *WOMAC Steifheit* aus RCT mit dem zu bewertenden Arzneimittel (Subgruppe *6MWT*)

UX023-CL303 Population (N)	Patienten n (%)	OR <sup>a,b</sup> [95% KI]	p-Wert <sup>a</sup>	RR <sup>b,c</sup> [95% KI]	ARD <sup>b,c</sup> [95% KI]	p-Wert <sup>d</sup>
<b><i>6MWT ≤ 365 m</i></b>						
Burosumab (37)	12 (32,43)	13,440 [1,564; 115,465]	0,0187	9,405 [1,297; 68,205]	0,290 [0,125; 0,455]	0,0040
Placebo (29)	1 (3,45)					
<b><i>6MWT &gt; 365 m</i></b>						
Burosumab (30)	8 (26,67)	1,091 [0,353; 3,369]	0,8780	1,067 [0,470; 2,422]	0,017 [-0,196; 0,229]	1,0000
Placebo (36)	9 (25,00)					

N: Anzahl der Patienten in der Studie; n: Anzahl der Responder; %: Anteil von N; [95% KI]: 95%-iges Konfidenzintervall; OR: *Odds Ratio*, RR: *Risk Ratio* (relatives Risiko); ARD: *Adjusted Risk Difference* (Absolute Risikodifferenz), *normalised units*

Ein RR bzw. OR > 1 zeigt einen Vorteil für Burosumab an und ein RR bzw. OR < 1 einen Vorteil für Placebo.

a: Das logistische Regressionsmodell beinhaltet die Behandlungsgruppe als unabhängige Variable.

b: Berechnung der Effektschätzer (OR, RR und ARD) erfolgte *post-hoc*.

c: Das RR und die ARD wurden unadjustiert aus den zum jeweiligen Endpunkt gehörenden Vierfeldertafeln berechnet. Die zugehörigen 95% KI wurden mittels Normalverteilungsapproximation berechnet.

d: Der p-Wert wurde mittels Fishers exaktem Test berechnet.

Für den Endpunkt *WOMAC Steifheit* kann eine potenzielle Effektmodifikation in der Subgruppe *6MWT zu Baseline* ( $\leq 365 m$  vs.  $> 365 m$ ) beobachtet werden. Bei einem p-Wert im Interaktionstest von  $p = 0,0278$  (Woche 24) profitieren Studienteilnehmer mit der Ausprägung *6MWT zu Baseline*  $\leq 365 m$  im Vergleich zur Ausprägung *6MWT zu Baseline*  $> 365 m$  mehr von einer Behandlung mit Burosumab. Während in der Ausprägung *6MWT zu Baseline*  $\leq 365 m$  ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen vorliegt (Woche 24: RR <sub>$\leq 365$</sub>  [95% KI]: 9,405; [1,297; 68,205];  $p = 0,0040$ ), ist der Unterschied zwischen den Behandlungsgruppen für die Ausprägung *6MWT zu Baseline*  $> 365 m$  (W<sub>24</sub>: RR <sub>$> 365$</sub>  [95% KI]: 1,067; [0,470; 2,422];  $p = 1,0000$ ) nicht signifikant (Tabelle 4-68). Aufgrund dessen erfolgt die Ableitung des Zusatznutzens ausschließlich aus den Ergebnissen für die Gesamtpopulation.

### ***Fazit der Subgruppenanalysen***

Die beobachteten Effektschätzer waren in den meisten Subgruppenanalysen für beide Ausprägungen eines Subgruppenmerkmals gleichgerichtet. Waren die Effektschätzer entgegengesetzt, so war stets lediglich der Effektschätzer einer der beiden Merkmalsausprägungen einer Subgruppe statistisch signifikant.

Generell ist zu beachten, dass bei einer Testung von zahlreichen statistischen Hypothesen eine gewisse Anzahl der durchgeführten Tests aufgrund der angenommenen Irrtumswahrscheinlichkeit zu einem falsch positiven Ergebnis führen. Die beobachteten positiven Interaktionstests werden als zufällige Effekte bewertet und nicht als Nachweis einer tatsächlichen Effektmodifikation. Die Ableitung des Zusatznutzens erfolgt daher für alle Endpunkte anhand der Gesamtpopulation.

#### **4.3.1.4 Liste der eingeschlossenen Studien - RCT**

*Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).*

Tabelle 4-69: Liste der eingeschlossenen Studien - RCT

<b>Studie</b>	<b>Datenquellen</b>
UX023-CL303 RCT	Studienbericht [5]
	Studienprotokoll [50]
	SAP [51]
	Studienregistereinträge [10-12]
	Studienpublikation [4, 23, 47]
	Dokumente von der Internetseite des G-BA [1, 37, 46, 48, 49]

### **4.3.2 Weitere Unterlagen**

#### **4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien**

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn indirekte Vergleiche als Nachweis für einen Zusatznutzen herangezogen werden sollen. Das ist dann möglich, wenn keine direkten Vergleichsstudien für das zu bewertende Arzneimittel gegenüber der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen.

**4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche**

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu Studien für indirekte Vergleiche. **Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.).** Benennen Sie sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie

- Studien des pharmazeutischen Unternehmers
- Studien aus der bibliografischen Literaturrecherche
- Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken
- Studien aus der Suche auf der Internetseite des G-BA
- Resultierender Studienpool aus den einzelnen Suchschritten

Nicht zutreffend.

**4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche**

Charakterisieren Sie nachfolgend die Studien, die für indirekte Vergleiche identifiziert wurden und bewerten Sie darüber hinaus deren Ähnlichkeit. Begründen Sie darauf basierend den Ein- bzw. Ausschluss von Studien für die von Ihnen durchgeführten indirekten Vergleiche. Bewerten Sie das Verzerrungspotenzial der für indirekte Vergleiche herangezogenen Studien. **Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.**

Nicht zutreffend.

**4.3.2.1.3 Ergebnisse aus indirekten Vergleichen**

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus indirekten Vergleichen beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-70: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche

Studie	<Mortalität>	<Gesundheits- bezogene Lebensqualität>	<Endpunkt>	<Endpunkt>	<Endpunkt>
<Studie 1>	nein	ja	ja	ja	nein

Nicht zutreffend.

**4.3.2.1.3.1 <Endpunkt xxx> – indirekte Vergleiche aus RCT**

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in drei Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. **Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.1.1.1.1.**

*Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche. Beispielhaft wäre folgende Darstellung denkbar:*

Tabelle 4-71: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Studie	Intervention	<Vergleichs-therapie 1>	<Vergleichs-therapie 2>	<Vergleichs-therapie 3>
1	<Studie 1>	•		•	•
2	<Studie 2> <Studie 3>	• •		• •	
1	<Studie 4>		•	•	•
etc.	etc.	etc.	etc.		

Nicht zutreffend.

*Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.*

Nicht zutreffend.

*Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.*

Tabelle 4-72: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Nicht zutreffend.

*Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.*

*Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.*

Tabelle 4-73: Bewertung des Verzerrungspotenzials für &lt;Endpunkt xxx&gt; in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>

Nicht zutreffend.

*Begründen Sie für jede Studie die abschließende Einschätzung.*

Nicht zutreffend.

*Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.*

Tabelle 4-74: Ergebnisse für &lt;Endpunkt xxx&gt; aus RCT für indirekte Vergleiche

Studie	Tabellarische Präsentation in geeigneter Form (Anforderungen siehe Erläuterung in Abschnitt 4.1.1.1.1)
<Studie 1>	

Nicht zutreffend.

*Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an der üblichen Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:*

- *Homogenität der Ergebnisse: Stellen Sie die Ergebnisse der paarweisen Meta-Analysen dar. Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere inkonsistente Ergebnisse.*

*Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.*

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt für den ein indirekter Vergleich vorgenommen wird fortlaufend in einem eigenen Abschnitt dar.

#### **4.3.2.1.3.2 Subgruppenanalysen – indirekte Vergleiche aus RCT**

*Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen auf Basis indirekter Vergleiche aus RCT. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.*

Nicht zutreffend.

#### **4.3.2.1.4 Liste der eingeschlossenen Studien – indirekte Vergleiche aus RCT**

*Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).*

Nicht zutreffend.

#### **4.3.2.2 Nicht randomisierte vergleichende Studien**

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn nicht randomisierte vergleichende Studien als Nachweis für einen Zusatznutzen herangezogen werden sollen.

##### **4.3.2.2.1 Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien**

*Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie*

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken*
- *Studien aus der Suche auf der G-BA Internetseite*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

Nicht zutreffend.

**4.3.2.2.2 Charakteristika der nicht randomisierten vergleichenden Studien**

*Charakterisieren Sie nachfolgend die nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.*

*Beschreiben Sie die Verzerrungsaspekte der nicht randomisierten vergleichenden Studie auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Beschreibung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.*

*Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.*

Tabelle 4-75: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
			Patient	Behandelnde Personen		
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>

*Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Studienebene.*

Nicht zutreffend.

#### 4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus nicht randomisierten vergleichenden Studien beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-76: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien

Studie	<Mortalität>	<Gesundheits- bezogene Lebensqualität>	<Endpunkt>	<Endpunkt>	<Endpunkt>
<Studie 1>	nein	ja	ja	ja	nein

##### 4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-77: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Beschreiben Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-78: Verzerrungsaspekte für &lt;Endpunkt xxx&gt; – nicht randomisierte vergleichende Studien

Studie	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Endpunktebene.

Nicht zutreffend.

Stellen Sie die Ergebnisse der nicht randomisierten vergleichenden Studien gemäß den Anforderungen des TREND- bzw. des STROBE-Statements dar. Machen Sie dabei auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus nicht randomisierten vergleichenden Studien fortlaufend in einem eigenen Abschnitt dar.

#### 4.3.2.2.3.2 Subgruppenanalysen – nicht randomisierte vergleichende Studien

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus nicht randomisierten vergleichenden Studien. **Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.**

Nicht zutreffend.

#### 4.3.2.2.4 Liste der eingeschlossenen Studien – nicht randomisierte vergleichende Studien

Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).

Nicht zutreffend.

### 4.3.2.3 Weitere Untersuchungen

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn über die in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 genannten Studien hinausgehende Untersuchungen als Nachweis für einen Zusatznutzen herangezogen werden sollen.

#### 4.3.2.3.1 Ergebnis der Informationsbeschaffung – weitere Untersuchungen

*Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung nach Untersuchungen, die nicht in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 aufgeführt sind. **Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie***

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken*
- *Studien aus der Suche auf der G-BA Internetseite*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

Nicht zutreffend.

#### 4.3.2.3.2 Charakteristika der weiteren Untersuchungen

*Charakterisieren Sie nachfolgend die weiteren Untersuchungen und bewerten Sie deren Verzerrungsaspekte.*

*Ergebnisse nicht randomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.*

***Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.***

Nicht zutreffend.

#### 4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus weiteren Untersuchungen beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-79: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen

Studie	<Mortalität>	<Gesundheits- bezogene Lebensqualität>	<Endpunkt>	<Endpunkt>	<Endpunkt>
<Studie 1>	nein	ja	ja	ja	nein

#### 4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-80: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt. Ergebnisse nicht randomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

Nicht zutreffend.

Stellen Sie die Ergebnisse der weiteren Untersuchungen gemäß den jeweils gültigen Standards für die Berichterstattung dar. Begründen Sie dabei die Auswahl des Standards für die Berichterstattung. Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus weiteren Untersuchungen fortlaufend in einem eigenen Abschnitt dar.

#### 4.3.2.3.3.2 Subgruppenanalysen – weitere Untersuchungen

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus weiteren Untersuchungen. **Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.**

Nicht zutreffend.

#### **4.3.2.3.4 Liste der eingeschlossenen Studien – weitere Untersuchungen**

*Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).*

Nicht zutreffend.

## 4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens

### 4.4.1 Beurteilung der Aussagekraft der Nachweise

*Legen Sie für alle im Dossier eingereichten Unterlagen die Evidenzstufe dar. Beschreiben Sie zusammenfassend auf Basis der in den Abschnitten 4.3.1 und 4.3.2 präsentierten Ergebnisse die Aussagekraft der Nachweise für einen Zusatznutzen unter Berücksichtigung der Studienqualität, der Validität der herangezogenen Endpunkte sowie der Evidenzstufe.*

Bei Burosumab (Crysvita®) handelt es sich gemäß § 35a Abs. 1 Satz 11 SGB V um ein Arzneimittel zur Behandlung eines seltenen Leidens. Der *Orphan-Drug*-Status wird Arzneimitteln verliehen, die zur Behandlung einer seltenen chronischen oder lebensbedrohlichen Erkrankung eingesetzt werden, für die keine zufriedenstellende Therapie existiert. Nachweise gemäß § 35a Abs. 1 Satz 3 Nummer 2 und 3 SGB V über den medizinischen Nutzen und Zusatznutzen im Verhältnis zur zweckmäßigen Vergleichstherapie (zVT) müssen nicht erbracht werden. Lediglich das Ausmaß des Zusatznutzens ist für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, nachzuweisen (5. Kapitel § 12 Nr. 1 Satz 2 Verfo des G-BA). Übersteigt der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten zwölf Kalendermonaten einen Betrag von 50 Millionen Euro, hat der pharmazeutische Unternehmer gemäß § 35a Absatz 1 Satz 12 SGB V innerhalb von drei Monaten nach Aufforderung durch den G-BA Nachweise gemäß 5. Kapitel § 5 Absatz 1 bis 6 Verfo des G-BA zu übermitteln.

Der G-BA hat mit Schreiben vom 21.10.2021 die Kyowa Kirin GmbH darüber informiert, dass der Umsatz mit Burosumab (Crysvita®) zu Lasten der GKV einen Betrag von 50 Millionen Euro überschritten hat. Maßgeblicher Zeitpunkt für die Einreichung der Nachweise gemäß § 35a Abs. 1 Satz 12 SGB V ist damit der 01.02.2022. Im vorliegenden Dossier werden nun die erforderlichen Nachweise des Zusatznutzens von Burosumab gegenüber der zVT dargestellt.

Für die erneute Bewertung von Burosumab nach Überschreitung der Umsatzgrenze von 50 Millionen Euro werden die Studiendaten der randomisierten, doppelblinden, multizentrischen, Placebo-kontrollierten Studie UX023-CL303 herangezogen, die bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 bewertet wurden [1]. Gemäß § 5 Abs. 6 AM-NutzenV handelt es sich bei der Studie UX023-CL303 um eine Studie der Evidenzstufe 1b, dem höchsten Evidenzgrad für randomisierte, klinische Studien.

Um eine hohe Ergebnissicherheit mit niedrigem Verzerrungspotenzial zu gewährleisten, wurde ausschließlich die 24-wöchige doppelblinde Phase der Studie für das Ableiten des Zusatznutzens von Burosumab herangezogen.

In der Studie UX023-CL303 wurden die Patienten 1:1 mittels *Interactive Web Response System* (IWRS) auf die Behandlungsarme randomisiert. Ursprünglich sollte die Stratifizierung anhand der Region sowie dem durch den Patienten wahrgenommenen schlimmsten Schmerz (Frage 3 des BPI-Fragebogens) erfolgen. Aufgrund eines Fehlers im IWRS wurde für die Stratifizierung jedoch die Frage 5 des BPI-Fragebogens *BPI Durchschnittlicher Schmerz* verwendet. Da die

Studienteilnehmerrekrutierung bereits stattfand als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Der Pearson Korrelationskoeffizient  $r$  zwischen den beiden Kategorien (*BPI Schlimmster Schmerz* und *BPI Durchschnittlicher Schmerz*) betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Unterstützt wird dies ebenfalls durch die Patientencharakteristika, die die gute Vergleichbarkeit der Behandlungsarme belegen. Zusätzlich durchgeführte Sensitivitätsanalysen, die für den ursprünglich geplanten Stratifizierungsfaktor *BPI Schlimmster Schmerz* adjustieren, bestätigten zudem die Primäranalysen, sodass weiterhin von einer hohen Ergebnissicherheit ausgegangen werden kann.

Sowohl die Studienteilnehmer als auch das klinische Personal waren gegenüber der Zuteilung verblindet. Die Validität und Patientenrelevanz der Endpunkte sind in Abschnitt 4.2.5.2 ausführlich dargelegt. Es fand zu allen Zeitpunkten eine ergebnisunabhängige Berichtserstattung statt.

Zusammenfassend lässt sich somit feststellen, dass die Studie UX023-CL303 sowohl auf Studien- als auch auf Endpunktebene ein niedriges Verzerrungspotenzial aufweist und eine hohe Ergebnissicherheit während der gesamten doppelblinden Phase der Studie gegeben ist. Gemäß des Methodenpapiers des IQWiG ist die Aussagesicherheit der dargestellten Nachweise zum medizinischen Zusatznutzen von Burosumab bei der Behandlung von Erwachsenen mit XLH daher als **Hinweis** einzustufen [8].

#### **4.4.2 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß**

*Führen Sie die in den Abschnitten 4.3.1 und 4.3.2 beschriebenen Ergebnisse zum Zusatznutzen auf Ebene einzelner Endpunkte zusammen und leiten Sie ab, ob sich aus der Zusammenschau der Ergebnisse zu den einzelnen Endpunkten insgesamt ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zweckmäßigen Vergleichstherapie ergibt. Berücksichtigen Sie dabei auch die Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext. Liegt ein Zusatznutzen vor, beschreiben Sie worin der Zusatznutzen besteht.*

*Stellen Sie die Wahrscheinlichkeit des Zusatznutzens dar, d. h., beschreiben und begründen Sie unter Berücksichtigung der in Abschnitt 4.4.1 dargelegten Aussagekraft der Nachweise die Ergebnissicherheit der Aussage zum Zusatznutzen.*

*Beschreiben Sie außerdem das Ausmaß des Zusatznutzens unter Verwendung folgender Kategorisierung (in der Definition gemäß AM-NutzenV):*

- *erheblicher Zusatznutzen*
- *beträchtlicher Zusatznutzen*
- *geringer Zusatznutzen*
- *nicht quantifizierbarer Zusatznutzen*

- kein Zusatznutzen belegbar
- der Nutzen des zu bewertenden Arzneimittels ist geringer als der Nutzen der zweckmäßigen Vergleichstherapie

Berücksichtigen Sie bei den Aussagen zum Zusatznutzen ggf. nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen.

Für die Neubewertung nach Überschreitung der Umsatzgrenze von 50 Millionen Euro werden die Daten der Studie UX023-CL303 abermals im Dossier aufgeführt und für die Ableitung des Zusatznutzens von Burosumab zur Behandlung von Erwachsenen mit XLH herangezogen.

In der nachfolgenden Tabelle 4-81 ist eine Übersicht der Ergebnisse der patientenrelevanten Endpunkte der Kategorie *Morbidität* in der Studie UX023-CL303 dargestellt.

Tabelle 4-81 Zusammenfassung der Ergebnisse der Kategorie *Morbidität* zum medizinischen Zusatznutzen von Burosumab

Patientenrelevante Endpunkte	UX023-CL303 Burosumab (N = 68) vs. Placebo (N = 66) Behandlungsdifferenz (B-P) Burosumab vs. Placebo Woche 24
<b>Serumphosphat</b>	
<i>Serumphosphat (mg/dl)</i>	
LsMean <sup>a</sup>	0,40
[95% KI]	[0,28; 0,51]
p-Wert	< 0,0001
Hedges`g <sup>b</sup> [95% KI]	H`g = 0,922 [0,566; 1,278]
<i>Serumphosphat Prozentsatz erreichter Normwert (≥ 2,5 mg/dl)</i>	
Burosumab (n (%)) vs. Placebo (n (%))	63 (92,6) vs. 5 (7,6)
RR <sup>a</sup>	12,229
[95% KI]	[5,252; 28,479]
p-Wert	< 0,0001
<b>Gefähigkeit</b>	
<i>6MWT – Messung in Meter<sup>c</sup></i>	
LsMean <sup>a</sup>	19,83
[95% KI]	[4,19; 35,47]
p-Wert	0,0129
Hedges`g <sup>b</sup> [95% KI]	H`g = 0,424 [0,075; 0,773]

<b>Patientenrelevante Endpunkte</b>	<b>UX023-CL303 Burosumab (N = 68) vs. Placebo (N = 66) Behandlungsdifferenz (B-P) Burosumab vs. Placebo Woche 24</b>
<i>6MWT Gesamtrate – Prozentsatz<sup>c</sup></i>	
LsMean <sup>a</sup>	3,19
[95% KI]	[0,96; 5,41]
p-Wert	0,0050
Hedges`g <sup>b</sup> [95% KI]	H`g = 0,460 [0,110; 0,809]
<b>Steifheit und Physische Funktion</b>	
<i>WOMAC Steifheit</i>	
LsMean <sup>a</sup>	-8,31
[95% KI]	[-14,68, -1,94]
p-Wert	0,0106
Hedges`g <sup>b</sup> [95% KI]	H`g = -0,420 [-0,765; -0,075]
<i>WOMAC Steifheit ≥ 15 nu</i>	
Burosumab (n (%)) vs. Placebo (n (%))	21 (30,9) vs. 10 (15,2)
RR <sup>a</sup>	2,038
[95% KI]	[1,040; 3,994]
p-Wert	0,0403
<i>WOMAC Physische Funktion</i>	
LsMean <sup>a</sup>	-4,90
[95% KI]	[-9,76, -0,05]
p-Wert	0,0478
Hedges`g <sup>b</sup> [95% KI]	H`g = -0,389 [-0,735; -0,043]
<i>WOMAC Physische Funktion ≥ 15 nu</i>	
Burosumab (n (%)) vs. Placebo (n (%))	14 (20,6) vs. 11 (16,7)
RR <sup>a</sup>	1,235
[95% KI]	[0,605; 2,521]
p-Wert	0,6591
<b>Schmerzerhebung</b>	
<i>BPI Schlimmster Schmerz</i>	
LsMean <sup>a</sup>	-0,46
[95% KI]	[-1,00; 0,08]
p-Wert	0,0919
Hedges`g <sup>b</sup> [95% KI]	-
<i>BPI Schlimmster Schmerz ≥ 15 %</i>	
Burosumab (n (%)) vs. Placebo (n (%))	29 (42,6) vs. 23 (34,8)
RR <sup>a</sup>	1,224
[95% KI]	[0,796; 1,881]
p-Wert	0,3800

<b>Patientenrelevante Endpunkte</b>	<b>UX023-CL303 Burosumab (N = 68) vs. Placebo (N = 66) Behandlungsdifferenz (B-P) Burosumab vs. Placebo Woche 24</b>
<i>BPI Schlimmster Schmerz <math>\geq 30</math> %</i> Burosumab (n (%)) vs. Placebo (n (%)) RR <sup>a</sup> [95% KI] p-Wert	17 (25,0) vs. 12 (18,2) 1,375 [0,713; 2,652] 0,4036
<i>BPI Schmerzintensität</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,43 [-0,93; 0,07] 0,0940 -
<i>BPI Schmerzinterferenz</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,13 [-0,70; 0,44] 0,6527 -
<b>Fatigue</b>	
<i>BFI Schlimmste Fatigue</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,20 [0,40; -0,80] 0,5102 -
<i>BFI Globale Fatigue</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	0,13 [0,70, -0,44] 0,6572 -
<b>Erfassung von Therapieeffekten</b>	
<i>PGI-I</i> LsMean <sup>a</sup> [95% KI] p-Wert Hedges`g <sup>b</sup> [95% KI]	-0,29 [-0,74; 0,16] 0,2073 -
<p>(B-P): Behandlungsdifferenz betrachtet von Baseline zu Post-Baseline (Woche 24) der Burosumab-Gruppe mit Placebo, 6MWT: 6-Minuten-Gehdistanz Test, WOMAC: <i>Western Ontario and McMaster Universities Osteoarthritis</i>, BPI: <i>Brief Pain Inventory</i>, BFI: <i>Brief Fatigue Inventory</i>, PGI-I: <i>Patient Global Impression of Improvement</i>; [95% KI]:95%-Konfidenzintervall, RR: <i>Risk Ratio</i> (relatives Risiko)</p> <p>Ein RR &gt; 1 zeigt einen Vorteil für Burosumab an und ein RR &lt; 1 einen Vorteil für Placebo.</p> <p>a: LsMean/RR: Eine detaillierte Beschreibung des statistischen Modells ist in Abschnitt 4.2.5.2 und 4.3.1.3 zu finden.</p> <p>b: Hedges`g-Berechnung erfolgte nur falls p &lt; 0,05.</p> <p>c: 6MWT: Aufgrund eines Kalkulationsfehlers durch das Studienzentrum zu Beginn der Studie, wurde zunächst ein fehlerhafter Baseline-Wert in die Datenbank eingetragen, der nach der Primäranalyse zu Woche 24 korrigiert wurde. Die korrigierte Analyse berücksichtigt ausschließlich die tatsächlich zurückgelegte Distanz des Studienteilnehmers nach der</p>	

<b>Patientenrelevante Endpunkte</b>	<b>UX023-CL303</b> <b>Burosumab (N = 68) vs. Placebo (N = 66)</b> <b>Behandlungsdifferenz (B-P)</b> <b>Burosumab vs. Placebo</b> <b>Woche 24</b>
Korrektur des Kalkulationsfehlers. Diese korrigierte Analyse wurde bereits im Nutzenbewertungsverfahren 2020-11-01-D-588 vom G-BA anerkannt und zur Ableitung des Zusatznutzens herangezogen [1].	

**Mortalität**

Während der 24-wöchigen doppelblinden Phase der Studie UX023-CL303 traten keine Todesfälle in den Behandlungsgruppen (Burosumab-Gruppe und Placebo-Gruppe) auf.

**Morbidität*****Serumphosphat***

*Serumphosphat (mg/dl) und Serumphosphat – Prozentsatz erreichter Normwert ( $\geq 2,5$  mg/dl)*

Das primäre Therapieziel der XLH, eine Normalisierung der Serumphosphat-Konzentration (LLN:  $\geq 2,5$  mg/dl), konnte unter der bisher einzigen verfügbaren Therapieoption, einer Substitutionstherapie bestehend aus oralem Phosphat und aktivem Vitamin D, nicht erreicht werden. Dies wird auch anhand der Charakteristika der Studienpopulation der Studie UX023-CL303 ersichtlich, bei der die Serumphosphat-Konzentration zu Baseline, trotz jahrelanger Substitutionstherapie, im Durchschnitt unterhalb des unteren Normbereichs von 2,5 mg/dl lag (Burosumab: 2,03 mg/dl, Placebo: 1,92 mg/dl). Der bestehende chronische Phosphatverlust führt zu zahlreichen schweren Folgeerkrankungen. Hierzu zählen unter anderem Skelettdeformationen, Enthesopathien, Osteomalazie, Arthrose und Pseudofrakturen. Bedingt durch die daraus resultierende verminderte Mobilität, den stark ausgeprägten Schmerzen sowie häufig teils schweren Operationen sind diese XLH-Patienten mit erheblichen Beeinträchtigungen im Alltag konfrontiert.

Die Studie UX023-CL303 zeigt, dass durch Burosumab erstmalig eine Therapieoption für Erwachsene mit XLH zur Verfügung steht, durch die bereits innerhalb einer kurzen Zeit die Serumphosphat-Homöostase wiederhergestellt werden kann (Woche 12: Burosumab: MW: 2,69 (mg/dl); Placebo: MW: 2,06 (mg/dl)). Der positive Effekt auf die Serumphosphat-Konzentration bleibt auch langfristig zu Woche 24 bestehen. Die Behandlungsdifferenz belegt hierbei den sowohl statistisch signifikanten als auch klinisch relevanten Vorteil der Burosumab-Behandlung (LsMeanW24: 0,40 (mg/dl);  $p < 0,0001$ ,  $H^*g = 0,922$  [0,566; 1,278]) (Tabelle 4-81).

Das positive Ergebnis hinsichtlich des Serumphosphat-Homöostase wird auch durch die Responderanalyse bestätigt. Während in der Burosumab-Gruppe zu Woche 24 92,6 % der Studienteilnehmer eine Serumphosphat-Konzentration im unteren Normbereich (LLN:  $\geq 2,5$  mg/dl) erlangen, ist dies nur bei 7,6 % der Studienteilnehmer der Placebo-Gruppe der Fall (RR: 12,229;  $p < 0,0001$ ) (Tabelle 4-81).

Im Gegensatz zu der Substitutionstherapie aus oralem Phosphat und aktivem Vitamin D steht mit Burosumab somit erstmalig eine **wirksame** und **zielgerichtete** Therapieoption zur Verfügung, die direkt in den Pathomechanismus der XLH-Erkrankung eingreift und die **Serumphosphat-Homöostase** von XLH-Patienten **langfristig wiederherstellt**.

### ***Gehfähigkeit***

Die Gehfähigkeit von Erwachsenen mit XLH ist auf vielfältige Weise beeinträchtigt. Neben den Skelettdeformationen in den unteren Extremitäten, der Steifheit und der verringerten Muskelkraft führt die früheinsetzende Arthrose zu einer eingeschränkten Mobilität. In der Studie UX023-CL303 wurde der Einfluss der Burosumab-Behandlung auf die Gehfähigkeit anhand des 6MWT evaluiert. Hierbei wurde die durchschnittliche Gesamtstrecke in Metern, die innerhalb von 6 Minuten zurückgelegt werden konnte (absoluter Wert) gemessen. Zusätzlich wurde die innerhalb von 6 Minuten zurückgelegte Strecke als Prozentsatz der vorhergesagten Werte auf Grundlage normativer Daten ermittelt (relativer Wert).

Obwohl die Studienteilnehmer der Burosumab-Gruppe (MW: 356,78 m) zu Baseline eine stärkere Beeinträchtigung bezüglich der Gehfähigkeit als die Placebo-Gruppe (MW: 367,42 m) aufwiesen, konnten diese bereits nach 12 Wochen eine größere Distanz innerhalb von 6 min zurücklegen (Burosumab: MW: 373,97 m; Placebo: MW: 370,92 m). Die Gehfähigkeit steigerte sich weiterhin bis zu Woche 24, jedoch ausschließlich in der Burosumab-Gruppe (Burosumab: LsMean: 18,19 m; Placebo: LsMean: -1,65 m). Übereinstimmend hiermit zeigt auch der relative 6MWT zu Woche 24 eine Zunahme der vorausgesagten Gehdistanz in der Burosumab-Gruppe (Woche 24: LsMean: 2,56 %), während sich die Placebo-Gruppe (Woche 24: LsMean: -0,63 Prozentpunkte) im Vergleich zu Baseline verschlechterte. Die Behandlungsdifferenz zu Woche 24 ist sowohl für den absoluten (LsMean: 19,83 m;  $p = 0,0129$ ) als auch den relativen 6MWT (LsMean: 3,19 Prozentpunkte;  $p = 0,005$ ) statistisch signifikant und belegt den positiven Einfluss der Burosumab-Therapie auf die Mobilität der XLH-Patienten (Tabelle 4-81).

Die beobachtete signifikante Steigerung der Gehfähigkeit lässt sich physiologisch am wahrscheinlichsten auf eine Verbesserung der Muskelfunktion zurückführen. Diese ist bei XLH-Patienten aufgrund der chronischen Hypophosphatämie beeinträchtigt und mit Steifheit, eingeschränkter Mobilität sowie Gleichgewichts- und Koordinationsschwierigkeiten assoziiert.

### ***Steifheit und Physische Funktion:***

#### *WOMAC Steifheit und WOMAC Physische Funktion*

In der Studie UX023-CL303 wurde das Ausmaß der Steifheit und physischen Funktion anhand des patientenberichteten Fragebogens *WOMAC* evaluiert. Im Vergleich zu den bei Patienten mit Arthrose gewöhnlich festgestellten Werten wies die Studienpopulation zu Baseline ein stärkeres Ausmaß an Steifheit auf (Burosumab: MW: 64,71; Placebo: MW: 61,36), wodurch die ausgeprägte rheumatische Manifestation der XLH-Erkrankung verdeutlicht wird. Zu Woche 12 zeigt sich dagegen bereits eine deutliche Verbesserung zugunsten von Burosumab. Die statistisch signifikante Behandlungsdifferenz zu Woche 24 (LsMean: -8,31;  $p = 0,0106$ ) unterstreicht hierbei den langanhaltenden positiven Effekt der Burosumab-Therapie auf die

Steifheit von XLH-Patienten. Der Unterschied hinsichtlich des Anteils an Studienteilnehmern mit einer als klinisch relevant eingestuften Verbesserung der *Steifheit* (Abnahme um  $\geq 15$  nu) erreichte zu Woche 24 ebenfalls eine statistische Signifikanz zugunsten von Burosumab (Burosumab: 30,9 %; Placebo: 15,2 %; RR: 2,038;  $p = 0,0403$ ). Die Wirkung von Burosumab zeigt sich auch anhand der Domäne *WOMAC Physische Funktion*, bei der zu Woche 24 ein statistisch signifikanter Behandlungsunterschied zugunsten von Burosumab vorliegt (LsMean: -4,90;  $p = 0,0478$ ). Der Unterschied hinsichtlich des Anteils an Studienteilnehmern mit einer als klinisch relevant eingestuften Verbesserung der *physischen Funktion* ( $\geq 15$  nu) erreicht zwar keine statistische Signifikanz, jedoch ist auch hier ein Trend zu einem größeren Anteil an Respondern in der Burosumab-Gruppe (Burosumab: 20,6 %; Placebo: 16,7 %) feststellbar (Tabelle 4-81).

Die in der Studie UX023-CL303 demonstrierte signifikante Abnahme der Steifheit und signifikante Verbesserung der physischen Funktion deuten darauf hin, dass durch Burosumab nicht nur die Serumphosphat-Homöostase wiederhergestellt, sondern auch der ATP-Metabolismus normalisiert werden, wodurch eine Entspannung der Sehnen und Muskulatur möglich wird.

### **Schmerzerhebung**

Die durch XLH verursachten Schmerzen wurden anhand des BPI-Fragebogens erhoben, der sowohl den schlimmsten Schmerz, die Schmerzinterferenz als auch die Schmerzintensität misst. Der *BPI Schlimmste Schmerz* wurde zu Baseline durchschnittlich mit einem Score von 6,8 auf einer Skala von 0 – 10 angegeben und spiegelt das Ausmaß der starken chronischen Schmerzen wider, die XLH-Patienten aufgrund der Skelettdeformationen, der Arthrose, den Enthesopathien und der Osteomalazie ausgesetzt sind.

Trotz fehlender statistischer Signifikanz zeigt sich für die Domäne *BPI Schlimmster Schmerz* in der Studie UX023-CL303 zu Woche 24 eine deutliche Tendenz zu einer größeren Reduktion des *BPI Schlimmsten Schmerzes* zugunsten von Burosumab (LsMean: -0,46;  $p = 0,0919$ ). Dies wird auch anhand des Anteils an Respondern deutlich, die eine als klinisch relevant eingestufte Verbesserung der Domäne *BPI Schlimmster Schmerz* zu Woche 24 erreichen. Während in der Burosumab-Gruppe 42,6 % der Studienteilnehmer eine Abnahme um mindestens 15 % (kleinster relevanter Unterschied) aufweisen, ist dies nur bei 34,8 % der Studienteilnehmer in der Placebo-Gruppe der Fall. Übereinstimmend hiermit ist ebenfalls der Anteil an Respondern, der eine bedeutsame Linderung (Abnahme um  $\geq 30$  %) der chronischen Schmerzen zu Woche 24 erreicht, in der Burosumab-Gruppe größer (Burosumab: 25 %; Placebo: 18,2 %). Die positive Wirkung von Burosumab zeigt sich auch anhand der Domäne *BPI Schmerzintensität* bei der ebenfalls zu Woche 24 ein deutlicher Trend zu einer stärkeren Abnahme (LsMean: -0,43;  $p = 0,0940$ ) von XLH-bedingten Schmerzen unter der Burosumab-Therapie vorhanden ist. Lediglich bei der Domäne *BPI Schmerzinterferenz* ist nur eine leichte Verbesserung zugunsten von Burosumab zu Woche 24 feststellbar (LsMean: -0,13;  $p = 0,6527$ ) (Tabelle 4-81).

Die Ergebnisse der unterschiedlichen Domänen des *BPI* zeigen übereinstimmend, dass XLH-Patienten unter Burosumab weniger Schmerzen haben als unter der zweckmäßigen Vergleichstherapie „*best supportive care*“.

### ***Fatigue***

In der Studie UX023-CL303 wurde die *Fatigue* anhand des patientenberichteten Fragebogens *BFI* evaluiert, bei dem sowohl die *BFI Schlimmste Fatigue* als auch die *BFI Globale Fatigue* anhand einer Skala von 0 – 10 gemessen wurde. Ein höherer Wert reflektiert hierbei eine stärkere Ausprägung der *Fatigue*.

Sowohl zu Beginn als auch am Ende der 24-wöchigen doppelblinden Phase der Studie wiesen die Burosumab-Gruppe und die Placebo-Gruppe ein ähnliches Ausmaß an *Fatigue* auf (Behandlungsdifferenz zu Woche 24: *BFI Schlimmste Fatigue*: LsMean: -0,20; p = 0,5102; *BFI Globale Fatigue*: LsMean: 0,13; p = 0,6572) (Tabelle 4-81).

### ***Erfassung von Therapieeffekten***

#### ***PGI-I***

Die Erfassung von Therapieeffekten erfolgte anhand des patientenberichteten Fragebogens *PGI-I*. Die kategorische Skala reicht hierbei von 1 – 7, wobei 7 als stark verschlechtert und 1 als stark verbessert definiert wird.

Zu Baseline wurde der Gesundheitszustand in der Burosumab- und in der Placebo-Gruppe als ähnlich beeinträchtigt wahrgenommen (Burosumab: MW: 3,18; Placebo: 3,00). Am Ende der doppelblinden Phase der Studie UX023-CL303 unterschieden sich die Behandlungsgruppen ebenfalls nicht signifikant (LsMean: -0,28; p = 0,2161) (Tabelle 4-81).

### ***Fazit Morbidität***

Mit **Burosumab** (Crysvita®), einem rekombinanten, vollständig humanen, monoklonalen Antikörper (AK, IgG1), steht erstmalig eine **zielgerichtete Behandlungsoption** zur Verfügung, die **langanhaltend** eine **Normalisierung** des **Phosphathaushaltes** bei Erwachsenen XLH-Patienten erzielt<sup>17</sup> [47]. Die **Wiederherstellung der Serumphosphat-Homöostase** hat direkte Auswirkungen auf das Ausmaß der Steifheit sowie Einschränkung der physischen Funktion und führt zu einer Linderung der stark ausgeprägten chronischen Knochen- und Gelenkschmerzen. Infolgedessen werden die Mobilität und die körperliche Aktivität der XLH-Patienten signifikant gesteigert und eine **bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens** erzielt.

---

<sup>17</sup> UX023-CL303: Über einen Zeitraum von 96 Wochen konnte ein langanhaltender Effekt durch nachhaltige Verbesserung der biochemischen Werte wie z. B. des Serumphosphats gemessen werden.

## Sicherheit

In der nachfolgenden Tabelle 4-82 ist eine Übersicht der *TEAE*, der *schweren TEAE*, der *schwerwiegenden TEAE* und der *TEAE von besonderem Interesse* dargestellt. Im Anschluss daran erfolgt eine Beurteilung des Ausmaßes des Zusatznutzens.

Tabelle 4-82 Zusammenfassung der Ergebnisse der Kategorie *Sicherheit* zum medizinischen Zusatznutzen von Burosumab gegenüber der zVT

<b>Sicherheit</b>	<b>Burosumab (N = 68) vs. Placebo (66) Woche 24</b>
<b>Gesamtrate</b>	
<i>Jegliche TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	64 (94,1) vs. 61 (92,4)
RR <sup>a,b</sup>	1,018
[95% KI]	[0,930; 1,115]
p-Wert	0,7425
<i>Schwerwiegende TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	2 (2,9) vs. 1 (1,5)
RR <sup>a,b</sup>	1,941
[95% KI]	[0,180; 20,900]
p-Wert	1,000
<i>Schwere TEAE Grad 3 – 4</i>	
Burosumab (n (%)) vs. Placebo (n (%))	8 (11,8) vs. 8 (12,1)
RR <sup>a,b</sup>	0,971
[95% KI]	[0,387; 2,435]
p-Wert	1,000
<i>Therapieabbruch durch TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	0 (0) vs. 0 (0)
RR <sup>a,b</sup>	-
[95% KI]	-
p-Wert	-
<i>Tod durch TEAE</i>	
Burosumab (n (%)) vs. Placebo (n (%))	0 (0) vs. 0 (0)
RR <sup>a,b</sup>	-
[95% KI]	-
p-Wert	-
<b>TEAE von besonderem Interesse</b>	
<i>Reaktionen an der Einstichstelle</i>	
Burosumab (n (%)) vs. Placebo (n (%))	8 (11,8) vs. 8 (12,1)
RR <sup>a,b</sup>	0,971
[95% KI]	[0,387; 2,435]
p-Wert	1,0000

Sicherheit	Burosumab (N = 68) vs. Placebo (66) Woche 24
<i>Hypersensitivität</i>	
Burosumab (n (%)) vs. Placebo (n (%))	4 (5,9) vs. 4 (6,1)
RR <sup>a,b</sup>	0,971
[95% KI]	[0,253, 3,721]
p-Wert	1,0000
<i>Hyperphosphatämie</i>	
Burosumab (n (%)) vs. Placebo (n (%))	4 (5,9) vs. 0 (0)
RR <sup>a,b</sup>	8,739
[95% KI]	[0,479, 159,198]
p-Wert	0,1195
<i>Restless-Legs-Syndrom</i>	
Burosumab (n (%)) vs. Placebo (n (%))	8 (11,8) vs. 6 (9,1)
RR <sup>a,b</sup>	1,294
[95% KI]	[0,475, 3,528]
p-Wert	0,7791
<p>N: Patientenzahl der entsprechend definierten Population; n: Anzahl der Patienten mit mindestens einem Ereignis in der entsprechenden Kategorie; %: Anteil von N; <i>Treatment emergent adverse event</i> (unerwünschtes Ereignis nach Beginn der Behandlung); [95% KI]: 95%-iges Konfidenzintervall; RR: <i>Risk Ratio</i> (relatives Risiko)</p> <p>Ein RR &lt; 1 zeigt einen Vorteil für Burosumab an und ein RR &gt; 1 einen Vorteil für Placebo.</p> <p>a: RR: Eine detaillierte Beschreibung des statistischen Modells ist in Abschnitt 4.2.5.2 und 4.3.1.3.1.7 zu finden.</p> <p>b: Treten keine Ereignisse auf, wird ein Korrekturwert von 0,5 zu den Zellen der Vierfeldertafel hinzugefügt.</p>	

Während der doppelblinden Phase der Studie UX023-CL303 wurde kein signifikantes Sicherheitsbedenken der Burosumab-Therapie festgestellt. Die meisten *TEAE* waren mild oder moderat und der Anteil an Studienteilnehmern mit mindestens einem *TEAE* war in den Behandlungsgruppen nahezu identisch (Burosumab: 94,1 %; Placebo: 92,4 %). Schwerwiegende *TEAE* traten nur bei 2 Studienteilnehmer der Burosumab-Gruppe und bei einem Studienteilnehmer der Placebo-Gruppe auf. *Schwere TEAE* der Grade 3 oder 4, Therapieabbrüche durch *TEAE* oder Todesfälle waren dagegen nicht zu verzeichnen (Tabelle 4-82).

Für fast alle *TEAE* auf Ebene der MedDRA SOC und PT (die bei  $\geq 10$  % der Patienten in mindestens einer Behandlungsgruppe aufgetreten sind) liegen ebenfalls keine statistisch signifikanten Unterschiede zwischen den Behandlungsgruppen vor (Tabelle 4-58). Es zeigte sich jedoch in der *SOC Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen* ein statistisch signifikanter Unterschied zugunsten von Burosumab hinsichtlich des *PT Arthralgie* zu Woche 24 (RR [95% KI]: 0,364 [0,152; 0,873]; p = 0,0198) (Tabelle 4-58). Des Weiteren zeigte sich in der *SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums* hinsichtlich des *PT Schmerzen im Oropharynx* ein signifikanter Unterschied zu Woche 24 zugunsten von Burosumab (RR [95% KI]: 0,139 [0,018; 1,096]; p = 0,0317) (Tabelle 4-58).

Für *TEAE von besonderem Interesse* liegt kein statistisch signifikanter Unterschied zwischen der Burosumab- und der Placebo-Gruppe vor (Tabelle 4-82).

### **Fazit Sicherheit**

Die Studiendaten der Studie UX023-CL303 belegen die gute Verträglichkeit von Burosumab bei Erwachsenen mit XLH. Die meisten *TEAE* waren mild oder moderat und traten nahezu gleich häufig in den Behandlungsgruppen auf. Die beobachteten Hyperphosphatämien (Grad 1 und 2), die lediglich bei 4 der 68 Studienteilnehmer der Burosumab-Gruppe feststellbar waren, ließen sich durch eine Dosisreduktion beheben. Schwere Nebenwirkungen wie eine Nephrokalzinose oder ein Hyperparathyreoidismus, die zu den bekannten Nebenwirkungen einer Phosphatsubstitution zählen, waren während der gesamten doppelblinden Phase der Studie nicht zu verzeichnen. Zudem führten keine *TEAE* zu einem Tod oder vorzeitigem Therapieabbruch von Studienteilnehmern. Insgesamt ist eine Behandlung mit Burosumab somit durch ein **ausgewogenes Nutzen-Risiko-Profil** gekennzeichnet.

### **Subgruppenanalysen**

In den Subgruppenanalysen ergaben sich für keinen Endpunkt bedeutsame Effektmodifikationen. Aufgrund dessen wird der Zusatznutzen ausschließlich von der Gesamtpopulation abgeleitet.

### **Übertragbarkeit auf den deutschen Versorgungskontext**

Die Mehrzahl der Studienteilnehmer aus der Studie UX023-CL303 stammte aus Studienzentren in den USA und den europäischen Ländern Frankreich, Italien, Irland und Großbritannien. Zudem stellt XLH eine genetisch bedingte Phosphatverlustkrankheit dar, die weitestgehend unabhängig von Umwelteinflüssen ist. Aufgrund dessen ist eine Übertragbarkeit auf den deutschen Versorgungskontext in der vorliegenden Studie gewährleistet [58]. Um den Versorgungskontext in Deutschland abzubilden wird die BurGER-Studie an einem Studienzentrum in Deutschland durchgeführt. Diese schließt Erwachsene mit XLH ( $\geq 18$  Jahre) ein. Ziel der Studie ist es, neben der Effektivität von Burosumab auf physiologische Parameter auch Effekte auf die Beweglichkeit und Aktivität im Alltag der XLH-Patienten im deutschen Versorgungskontext zu untersuchen. Die Studiendauer ist auf ein Jahr mit einer Zwischenauswertung nach 22 Wochen ausgelegt.

### **Gesamtfazit zum medizinischen Zusatznutzen von Burosumab**

Bei der XLH handelt es sich um eine seltene genetisch bedingte chronische Multisystemerkrankung mit einem komplexen Krankheitsbild. Das Hauptmerkmal der Erkrankung ist ein chronischer Phosphatverlust, der durch eine unzureichende Rückresorption von Phosphat über die proximalen Nierentubuli entsteht. Da Phosphat bei zahlreichen zellulären Stoffwechselwegen eine wichtige Schlüsselfunktion spielt, führt der Verlust zu schwerwiegenden und den Alltag stark beeinträchtigenden Folgeerkrankungen. Hierzu zählen Skelettdeformationen, Arthrose, Osteomalazie, Enthesopathien, Pseudofrakturen und Spinalkanalstenosen, die mit einer starken Steifheit sowie chronischen Knochen- und Gelenkschmerzen einhergehen [19, 20, 56, 65]. Die in der Kindheit erworbenen Deformationen der unteren Extremitäten tragen zudem zu erheblichen Schwierigkeiten beim Gehen und zu ausgeprägten Ganganomalien bei [66]. Neben den skelettalen und rheumatischen Beschwerden weisen XLH-Patienten außerdem ein erhöhtes Risiko für kardiovaskuläre Erkrankungen wie z. B. Bluthochdruck und linksventrikuläre Hypertrophie auf [56]. Bedingt durch die

verminderte Mobilität, die stark ausgeprägten Schmerzen sowie die häufigen teils schweren Operationen sind die XLH-Patienten mit erheblichen Beeinträchtigungen im Alltag und einer stark reduzierten Lebensqualität konfrontiert [19, 20, 57].

Bis zur Zulassung von Burosumab stand keine zielgerichtete und wirksame Behandlungsoption für Erwachsene mit XLH zur Verfügung, die direkt in den Pathomechanismus der FGF23-vermittelten XLH eingreift. Aufgrund fehlender Therapieoptionen versuchte man daher den chronischen Phosphatverlust durch eine Phosphatsubstitution zu kompensieren. Dieser historisch gewachsene Therapieansatz hat sich aus Fallberichten und Fallreihen abgeleitet, die Wirksamkeit wurde jedoch noch nie in einem belastbaren klinischen Entwicklungsprogramm bestätigt oder quantifiziert. Hervorzuheben ist zudem, dass unter einer Phosphatsubstitution sowohl der Phosphatverlust als auch die überschüssige biologische Aktivität von FGF23 bestehen bleiben. Das **primäre Therapieziel** bei der Behandlung der XLH, eine **Normalisierung der Serumphosphat-Konzentration (LLN: 2,5 mg/dl)** und eine damit verbundene ausreichende Versorgung des Körpers mit Phosphat, konnte deshalb bisher nicht erlangt werden. Trotz einer langjährigen Phosphatsubstitution ließ sich daher das Fortschreiten der Erkrankung nicht aufhalten.

Die Langzeittherapie mit Phosphat ist nicht nur wenig effizient, sondern auch mit der Ausbildung von schwerwiegenden therapieassoziierten Folgeerkrankungen assoziiert. Hier sind vor allem die Nephrolithiasis, Nephrokalzinosen sowie ein sekundärer oder tertiärer Hyperparathyreoidismus zu nennen. Bedingt wird dies durch die hohen Dosen an oralem Phosphat, die notwendig sind, um zumindest eine kurzzeitige Erhöhung der Serumphosphat-Konzentration zu bewirken. Die transiente Erhöhung der Serumphosphat-Konzentration geht mit einer Abnahme der Serumkalzium-Konzentration einher, wodurch die Synthese des Parathormons stimuliert wird und folglich die Ausbildung eines sekundären oder tertiären Hyperparathyreoidismus begünstigt wird. Das bereits durch die XLH-Erkrankung bestehende Risiko zur Ausbildung eines Hyperparathyreoidismus wird durch die Phosphatsubstitution somit zusätzlich verstärkt [56].

Aufgrund der unzureichenden Wirksamkeit und den schwerwiegenden therapieassoziierten Folgeerkrankungen wird die Phosphatsubstitution häufig im Erwachsenenalter abgebrochen. Da keine weiteren Therapieoptionen zu Verfügung stehen, erfolgt die Behandlung der XLH-Patienten anschließend lediglich im Sinne von „*best supportive care*“. Neben teils schweren Operationen beinhaltet dies zusätzlich eine symptomatische Therapie, um die anhaltenden Knochen- und Gelenkschmerzen, die Steifheit, die beeinträchtigte Muskelkraft und die Fatigue zu mildern. Der therapeutische Bedarf ist aufgrund der vielfältigen und schweren Symptomatik bei Erwachsenen mit XLH hoch.

Durch die Zulassung von Burosumab, der ersten zielgerichteten und wirksamen Therapie zur Behandlung von Erwachsenen mit XLH, kann diese therapeutische Lücke nun geschlossen werden. Burosumab bindet spezifisch das im Überschuss vorhandene FGF23, hemmt dessen biologische Aktivität und wirkt dem pathologischen Phosphatverlust entgegen. Burosumab setzt somit direkt an der Ursache der XLH-Erkrankung an und ermöglicht eine **rasche** und

**langanhaltende Normalisierung des Phosphathaushaltes<sup>18</sup>** [47]. Die **Wiederherstellung der Serumphosphat-Homöostase** hat direkte Auswirkungen auf das Ausmaß der Steifheit sowie die Einschränkung der physischen Funktion und führt zu einer Linderung der stark ausgeprägten chronischen Knochen- und Gelenkschmerzen. Infolgedessen werden die Mobilität und die körperliche Aktivität der XLH-Patienten signifikant gesteigert und eine **bisher nicht erreichte Verbesserung der Symptome der XLH-Erkrankung** erzielt. Darüber hinaus belegt die Studie UX023-CL303 das ausgewogene **Nutzen-Risiko-Profil** der Burosumab-Therapie. Die schweren therapieassoziierten Folgeerkrankungen, die häufig unter einer Phosphatsubstitution zu beobachten sind, treten unter einer Burosumab-Therapie nicht auf.

Die ausreichende Versorgung des Körpers mit Phosphat wirkt sich zudem positiv auf die Knochenmineralisierung und die Mikroarchitektur des Knochens aus und fördert das Heilen von (Pseudo-)Frakturen und eine Abnahme der Osteomalazie [4, 22, 47].

Langzeitdaten der Studie UX023-CL303 (über 48 und 96 Wochen) belegen, dass diese positiven Effekte auch dauerhaft unter der Burosumab-Therapie aufrechterhalten werden. XLH-Patienten, die nach 24 Wochen aus ethischen Gründen von der Placebo-Gruppe in die Burosumab-Gruppe wechselten, zeigten eine sofortige und anhaltende Verbesserung (über 72 Wochen) krankheitsrelevanter Parameter analog zu den XLH-Patienten, die bereits in der 24-wöchigen Placebo-kontrollierten Phase der Studie eine Burosumab-Behandlung erhielten. So werden in der Studie UX023-CL303 über den gesamten Zeitraum von zwei Jahren physiologische Serumphosphat-Konzentrationen erreicht, was zu einem langfristigen Rückgang von Knochen- und Gelenkschmerzen, Müdigkeit und Steifheit führt [23, 47].

Statistisch signifikante Verbesserungen gegenüber dem Ausgangswert bei den patientenberichteten Endpunkten (PRO) und der ambulanten Funktion wurden bereits früh beobachtet und hielten bis zu Woche 96 unter der Behandlung mit Burosumab an, wobei eine signifikante Abnahme von Schmerzen, Steifheit und Müdigkeit zu beobachten waren. Bei dem Endpunkt *WOMAC Steifheit* zeigte sich unter der Behandlung mit Burosumab bereits zu Woche 24 eine signifikante Verbesserung gegenüber dem Ausgangswert. Hervorzuheben ist zudem, dass unter einer Burosumab-Behandlung auch zu Woche 48 und Woche 96 signifikante Verbesserungen zu verzeichnen waren. Die bis zu Woche 96 beobachteten Verbesserungen der PRO unterstreichen den potenziellen Nutzen einer Langzeitbehandlung mit Burosumab. Trotz der zahlreichen Folgeerkrankungen und körperlichen Beeinträchtigungen werden klinisch bedeutsame Symptome gelindert.

Zusammengefasst zeigen diese Ergebnisse, dass die rasche und anhaltende Verbesserung der Serumphosphat-Homöostase durch die Behandlung mit Burosumab mit einer nachhaltigen Verringerung der erheblichen Krankheitslast einhergeht, die sich im Laufe des Lebens bei den Erwachsenen mit XLH zusammen mit muskuloskelettalen Schmerzen angesammelt hat.

---

<sup>18</sup> UX023-CL303: Über einen Zeitraum von 96 Wochen konnte ein langanhaltender Effekt durch nachhaltige Verbesserung der biochemischen Werte wie z. B. des Serumphosphats gemessen werden.

Durch die langanhaltende Wirksamkeit hat Burosumab somit das Potenzial, das weitere Fortschreiten der XLH-Erkrankung aufzuhalten und reversible Manifestationen zu modifizieren.

Unterstützt wird dies ebenfalls durch die BUR-02-Studie, in der über einen Zeitraum von 48 Wochen der Effekt von Burosumab auf patientenrelevante Endpunkte an einem Kollektiv aus 31 Patienten untersucht wurde, die zuvor bereits für zwei Jahre in der Studie UX023-CL303 mit Burosumab behandelt wurden [67].

Eine weitere unabhängige Studie, die BurGER-Studie (NCT04695860, einarmige, interventionelle Studie bei Erwachsenen mit XLH), wird derzeit an einem Studienzentrum in Deutschland durchgeführt. Ziel der Studie ist es, nicht nur die Wirksamkeit von Burosumab auf physiologische Parameter (primärer Endpunkt: Normalisierung der Serumphosphat-Konzentration) zu untersuchen, sondern auch die Auswirkungen auf die Mobilität und Aktivität im Alltag und der Lebensqualität von XLH-Patienten ( $\geq 18$  Jahre) im deutschen Versorgungskontext. Die Daten der ersten Zwischenanalyse zu Woche 22 bestätigen die bereits bekannte gute Wirksamkeit und Sicherheit von Burosumab bei der Behandlung von Erwachsenen mit XLH. Dies unterstreicht das Engagement von Kyowa Kirin weiter zu forschen, um den Betroffenen zu helfen und Fortschritte in der Patientenversorgung zu erzielen.

Aufgrund der raschen **Normalisierung des Phosphathaushaltes** und der damit verbundenen signifikanten **Verbesserung der Symptome der XLH-Erkrankung** (Einschränkung der Gehfähigkeit und physischen Funktion, Steifheit) sowie dem **ausgewogenen Nutzen-Risiko-Profil** liegt unter Burosumab gemäß § 5 Abs. 7 Nr. 2 AM-NutzenV eine bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens vor. Grundlage dieser Bewertung ist die doppelblinde, randomisierte, kontrollierte multizentrische Studie UX023-CL303, die gemäß § 5 Abs. 6 AM-NutzenV den Evidenzgrad 1b aufweist. Insgesamt ergibt sich daher ein **Hinweis auf einen beträchtlichen Zusatznutzen** für den Wirkstoff Burosumab bei der Behandlung von Erwachsenen mit XLH.

#### **4.4.3 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht**

*Geben Sie auf Basis der in den Abschnitten 4.3.1 und 4.3.2 beschriebenen Ergebnisse und unter Berücksichtigung des in Abschnitt 4.4.2 dargelegten Zusatznutzens sowie dessen Wahrscheinlichkeit und Ausmaß in der nachfolgenden Tabelle an, für welche Patientengruppen ein therapeutisch bedeutsamer Zusatznutzen besteht. Benennen Sie das Ausmaß des Zusatznutzens in Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Fügen Sie für jede Patientengruppe mit therapeutisch bedeutsamem Zusatznutzen eine neue Zeile ein.*

Tabelle 4-83: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
Erwachsene mit XLH, die innerhalb eines Jahres nicht auf eine Phosphatsubstitution ansprechen	Beträchtlicher Zusatznutzen

#### 4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte

##### 4.5.1 Begründung für die Vorlage indirekter Vergleiche

*Sofern mit dem Dossier indirekte Vergleiche (Abschnitt 4.3.2.1) eingereicht wurden, begründen Sie dies. Begründen Sie dabei auch, warum sich die ausgewählten Studien jeweils für einen indirekten Vergleich gegenüber dem zu bewertenden Arzneimittel und damit für den Nachweis eines Zusatznutzens durch indirekten Vergleich eignen.*

Nicht zutreffend.

##### 4.5.2 Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen

*Sofern mit dem Dossier nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) oder weitere Untersuchungen (Abschnitt 4.3.2.3) eingereicht wurden, nennen Sie die Gründe, nach denen es unmöglich oder unangemessen ist, zu den in diesen Studien bzw. Untersuchungen behandelten Fragestellungen Studien höchster Evidenzstufe (randomisierte klinische Studien) durchzuführen oder zu fordern.*

Nicht zutreffend.

##### 4.5.3 Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen

*Falls aus Ihrer Sicht valide Daten zu patientenrelevanten Endpunkten zum Zeitpunkt der Bewertung noch nicht vorliegen können, begründen Sie dies.*

Nicht zutreffend.

##### 4.5.4 Verwendung von Surrogatendpunkten

Die Verwendung von Surrogatendpunkten bedarf einer Begründung (siehe Abschnitt 4.5.3). Zusätzlich soll dargelegt werden, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.

Eine Validierung von Surrogatendpunkten bedarf in der Regel einer Meta-Analyse von Studien, in denen sowohl Effekte auf den Surrogatendpunkt als auch Effekte auf den interessierenden

patientenrelevanten Endpunkt untersucht wurden (Burzykowski 2005<sup>19</sup>, Molenberghs 2010<sup>20</sup>). Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrundeliegende Anwendungsgebiet und das zu bewertende Arzneimittel sowie die Vergleichstherapie erlauben.

Eine Möglichkeit der Verwendung von Surrogatendpunkten ohne abschließende Validierung stellt die Anwendung des Konzepts eines sogenannten Surrogate-Threshold-Effekts (STE) (Burzykowski 2006<sup>21</sup>) dar. Daneben besteht die Möglichkeit einer Surrogatvalidierung in der quantitativen Betrachtung geeigneter Korrelationsmaße von Surrogatendpunkt und interessierendem patientenrelevanten Endpunkt („individuelle Ebene“) sowie von Effekten auf den Surrogatendpunkt und Effekten auf den interessierenden patientenrelevanten Endpunkt („Studienebene“). Dabei ist dann zu zeigen, dass die unteren Grenzen der entsprechenden 95%- Konfidenzintervalle für solche Korrelationsmaße ausreichend hoch sind. Die Anwendung alternativer Methoden zur Surrogatvalidierung (siehe Weir 2006<sup>22</sup>) soll ausreichend begründet werden, insbesondere dann, wenn als Datengrundlage nur eine einzige Studie verwendet werden soll.

Berichten Sie zu den Studien zur Validierung oder zur Begründung für die Verwendung von Surrogatendpunkten mindestens folgende Informationen:

- Patientenpopulation
- Intervention
- Kontrolle
- Datenherkunft
- verwendete Methodik
- entsprechende Ergebnisse (abhängig von der Methode)
- Untersuchungen zur Robustheit
- ggf. Untersuchungen zur Übertragbarkeit

*Sofern Sie im Dossier Ergebnisse zu Surrogatendpunkten eingereicht haben, benennen Sie die Gründe für die Verwendung von Surrogatendpunkten. Beschreiben Sie, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.*

Nicht zutreffend.

---

<sup>19</sup> Burzykowski T (Ed.): The evaluation of surrogate endpoints. New York: Springer; 2005.

<sup>20</sup> Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M: A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Stat Methods Med Res 2010; 19(3): 205-236.

<sup>21</sup> Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.

<sup>22</sup> Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

#### 4.6 Referenzliste

Listen Sie nachfolgend alle Quellen (z. B. Publikationen, Studienberichte, Studienregister-einträge), die Sie im vorliegenden Dokument angegeben haben (als fortlaufend nummerierte Liste). Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard). Geben Sie bei Fachinformationen immer den Stand des Dokuments an.

1. Gemeinsamer Bundesausschuss (G-BA) (2021): 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 (Neues Anwendungsgebiet: X-chromosomale Hypophosphatämie,  $\geq 18$  Jahre) [Zugriff: 09.11.2021]. URL: [https://www.g-ba.de/downloads/39-261-4784/2021-04-15\\_AM-RL-XII\\_Burosumab\\_D-588\\_BAnz.pdf](https://www.g-ba.de/downloads/39-261-4784/2021-04-15_AM-RL-XII_Burosumab_D-588_BAnz.pdf).
2. Kyowa Kirin Holdings B.V. (2018): CRYSVITA 10/20/30 mg Injektionslösung; Fachinformation. Stand: 05/2021 [Zugriff: 22.12.2021]. URL: <http://www.fachinfo.de>.
3. Deutsche Gesellschaft für Kinderendokrinologie und -diabetologie (DGKED) e.V. (2016): S1-Leitlinie 147-008: Hereditäre hypophosphatämische Rachitiden. Stand: 03/2016 [Zugriff: 15.12.2021]. URL: [https://www.awmf.org/uploads/tx\\_szleitlinien/174-0081\\_S1\\_Hereditaere\\_hypophosphataemische\\_Rachitiden\\_2016-05-abgelaufen.pdf](https://www.awmf.org/uploads/tx_szleitlinien/174-0081_S1_Hereditaere_hypophosphataemische_Rachitiden_2016-05-abgelaufen.pdf).
4. Insogna KL, Briot K, Imel EA, Kamenický P, Ruppe MD, Portale AA, et al. (2018): 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. *Journal of Bone and Mineral Research*; 33(8):1383-93.
5. Ultragenyx pharmaceutical Inc. (2019): UX023-CL303: Clinical Study Report - Week 96: 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).
6. Fuchs J, Rabenberg M, Scheidt-Nave C (2013): Prevalence of selected musculoskeletal conditions in Germany. *Bundesgesundheitsblatt*; 56:678-86.
7. Gemeinsamer Bundesausschuss (G-BA) (2020): Niederschrift (finale Fassung) zum Beratungsgespräch gemäß AM-NutzenV Beratungsanforderung 2020-B-003 Burosumab zur Behandlung der X-chromosomalen Hypophosphatämie.
8. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2022): IQWiG Allgemeine Methoden Version 6.1. [Zugriff: 24.01.2022]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
9. European Medicines Agency (EMA) (2014): Public summary of opinion on orphan designation - Recombinant human monoclonal IgG1 antibody for fibroblast growth factor 23 for the treatment of X-linked hypophosphatemia [Zugriff: 16.12.2021]. URL: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Orphan\\_designation/2014/11/WC500177183.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Orphan_designation/2014/11/WC500177183.pdf).
10. Ultragenyx Pharmaceutical Inc (2015): UX023-CL303 2014-005529-11 ( EudraCT Number ) - 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) - ClinicalTrials.gov (NCT02526160). Stand des Eintrags: 05.02.2021. [Zugriff: 02.11.2021]. URL: <https://ClinicalTrials.gov/ct2/show/record/NCT02526160>

11. Ultragenyx Pharmaceutical Inc (2015): UX023-CL303 - 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) - EU-CTR (2014-005529-11). [Zugriff: 02.11.2021]. URL: [https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract\\_number:2014-005529-11](https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2014-005529-11)
12. Ultragenyx Pharmaceutical Inc (2015): 2014-005529-11 - 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) - WHO ICTRP (NCT02526160). Stand des Eintrags: 01.12.2020. [Zugriff: 02.11.2021]. URL: <https://trialssearch.who.int/Trial2.aspx?TrialID=NCT02526160>
13. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2019): Ivacaftor (Kombination mit Tezacaftor/Ivacaftor; zystische Fibrose, ab 12 Jahre, F508delMutation, heterozygot) – Nutzenbewertung gemäß § 35a SGB V [Zugriff: 27.12.2021]. URL: [https://www.g-ba.de/downloads/92-975-3269/2019-09-01\\_Nutzenbewertung-IQWiG\\_Ivacaftor-D-477.pdf](https://www.g-ba.de/downloads/92-975-3269/2019-09-01_Nutzenbewertung-IQWiG_Ivacaftor-D-477.pdf).
14. Gemeinsamer Bundesausschuss (G-BA) (2015): Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Secukinumab [Zugriff: 27.12.2021]. URL: [https://www.g-ba.de/downloads/40-268-3450/2015-11-27\\_AM-RL-XII\\_Secukinumab\\_2015-06-01-D167\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-3450/2015-11-27_AM-RL-XII_Secukinumab_2015-06-01-D167_TrG.pdf).
15. Kang YE, Hong JH, Kim J, Joung KH, Kim HJ, Ku BJ, et al. (2014): A novel PHEX gene mutation in a patient with sporadic hypophosphatemic rickets. *Endocrinology and Metabolism*; 29(2):195-201.
16. Chalmers J, Conacher W, Gardner D, Scott P (1967): Osteomalacia—a common disease in elderly women. *The Journal of bone and joint surgery British volume*; 49(3):403-23.
17. Wong SS-L, Wilczynski NL, Haynes RB (2006): Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. *Journal of the Medical Library Association*; 94(4):451.
18. Carpenter TO, Imel EA, Holm IA, Jan de Beur SM, Insogna KL (2011): A clinician's guide to X-linked hypophosphatemia. *Journal of Bone and Mineral Research*; 26(7):1381-8.
19. Lo S, Lachmann R, Williams A, Piglowska N, Lloyd A (2020): Exploring the burden of X-linked hypophosphatemia: a European multi-country qualitative study. *Quality of Life Research*; 29(7):1883.
20. Ferizović N, Marshall J, Williams AE, Mughal MZ, Shaw N, Mak C, et al. (2020): Exploring the Burden of X-Linked Hypophosphataemia: An Opportunistic Qualitative Study of Patient Statements Generated During a Technology Appraisal. *Advances in Therapy*; 37(2):770-84.
21. Linglart A, Biosse-Duplan M, Briot K, Chaussain C, Esterle L, Guillaume-Czitrom S, et al. (2014): Therapeutic management of hypophosphatemic rickets from infancy to adulthood. *Endocrine connections*; 3(1):R13-30.
22. Insogna KL, Rauch F, Kamenický P, Ito N, Kubota T, Nakamura A, et al. (2019): Burosumab Improved Histomorphometric Measures of Osteomalacia in Adults with

- X-Linked Hypophosphatemia: A Phase 3, Single-Arm, International Trial. *Journal of Bone and Mineral Research*; 34(12):2183-91.
23. Portale AA, Carpenter TO, Brandi ML, Briot K, Cheong HI, Cohen-Solal M, et al. (2019): 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. *Calcified tissue international*; 105(3):271-84.
  24. Friedman E (2005): Consequences and management of hyperphosphatemia in patients with renal insufficiency. *Kidney International*; 67:S1-S7.
  25. Carpenter TO, Imel EA, Ruppe MD, Weber TJ, Klausner MA, Wooddell MM, et al. (2014): Randomized trial of the anti-FGF23 antibody KRN23 in X-linked hypophosphatemia. *The Journal of clinical investigation*; 124(4):1587-97.
  26. Theodore-Oklotka C, Bonner N, Spencer H, Arbuckle R, Chen C-Y, Skrinar A (2018): Qualitative research to explore the patient experience of X-linked hypophosphatemia and evaluate the suitability of the BPI-SF and WOMAC® as clinical trial end points. *Value in Health*; 21(8):973-83.
  27. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. (2013): OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis and cartilage*; 21(8):1042-52.
  28. American Thoracic Society (ATS) (2002): American Thoracic Society (ATS) Statement: Guidelines for the six-minute walk test. *Am J Respir Crit Care Med*; 166(1):111-7.
  29. Ultragenyx Pharmaceutical Inc (2015): Clinical Evaluator User Manual Protocol: UX023-CL303 A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study to Assess the Efficacy and Safety of KRN23 in Adults with X-linked Hypophosphatemia (XLH) MANUAL VERSION 1.1.
  30. Gibbons WJ, Fruchter N, Sloan S, Levy RD (2001): Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years. *Journal of Cardiopulmonary Rehabilitation*; 21(2):87-93.
  31. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S, et al. (2011): Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. *Journal of clinical epidemiology*; 64(2):182-90.
  32. Salaffi F, Leardini G, Canesi B, Mannoni A, Fioravanti A, Caporali Ro, et al. (2003): Reliability and validity of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index in Italian patients with osteoarthritis of the knee. *OsteoArthritis and Cartilage* 11:551-60.
  33. Baron G, Tubach F, Ravaud P, Logeart I, Dougados M (2007): Validation of a short form of the Western Ontario and McMaster Universities Osteoarthritis Index function subscale in hip and knee osteoarthritis. *Arthritis and rheumatism*; 57(4):633-8.
  34. Bellamy N (2012): WOMAC® Osteoarthritis Index User Guide X.
  35. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N (2000): Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *The Journal of rheumatology*; 27(11):2635-41.
  36. Ultragenyx pharmaceutical Inc. (2018): Patient-Reported Outcome Measures Clinical Outcome Assessment Dossier, v12.0.
  37. Gemeinsamer Bundesausschuss (G-BA) (2021): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene

- Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo  
Wirkstoff: Burosumab. [Zugriff: 09.11.2021]. URL: [https://www.g-ba.de/downloads/92-975-4127/2020-11-01\\_Nutzenbewertung-G-BA\\_Burosumab\\_D-588.pdf](https://www.g-ba.de/downloads/92-975-4127/2020-11-01_Nutzenbewertung-G-BA_Burosumab_D-588.pdf).
38. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS (2004): Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. *The Clinical journal of pain*; 20(5):309-18.
  39. De Lorenzo F, Hargreaves J, Kakkar V (1998): Phosphate diabetes in patients with chronic fatigue syndrome. *Postgraduate medical journal*; 74(870):229-32.
  40. Murphy SL, Lyden AK, Smith DM, Dong Q, Koliba JF (2010): The effect of a tailored activity pacing intervention on pain and fatigue for adults with osteoarthritis. *Am J Occup Ther*; 64(6):869-76.
  41. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. (1999): The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. *Cancer*; 85(5):1186-96.
  42. Aynehchi BB, Obourn C, Sundaram K, Bentsianov BL, Rosenfeld RM (2013): Validation of the Modified Brief Fatigue Inventory in head and neck cancer patients. *Otolaryngol Head Neck Surg*; 148(1):69-74.
  43. Wolfe F (2004): Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. *J Rheumatol*; 31(10):1896-902.
  44. Mease PJ, Spaeth M, Clauw DJ, Arnold LM, Bradley LA, Russell IJ, et al. (2011): Estimation of minimum clinically important difference for pain in fibromyalgia. *Arthritis care & research*; 63(6):821-6.
  45. Zhang J, Kai FY (1998): What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *Jama*; 280(19):1690-1.
  46. Gemeinsamer Bundesausschuss (G-BA) (2021): Tragende Gründe zum 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 (Neues Anwendungsgebiet: X chromosomale Hypophosphatämie, ≥ 18 Jahre). [Zugriff: 09.11.2021]. URL: [https://www.g-ba.de/downloads/40-268-7448/237ba1a53ca4b4e6f20dfdee49bf929e/2021-04-15\\_AM-RL-XII\\_Burosumab\\_D-588\\_TrG.pdf](https://www.g-ba.de/downloads/40-268-7448/237ba1a53ca4b4e6f20dfdee49bf929e/2021-04-15_AM-RL-XII_Burosumab_D-588_TrG.pdf).
  47. Briot K, Portale AA, Brandi ML, Carpenter TO, Cheong HI, Cohen-Solal M, et al. (2021): Burosumab treatment in adults with X-linked hypophosphataemia: 96-week patient-reported outcomes and ambulatory function from a randomised phase 3 trial and open-label extension. *RMD open*; 7(3):e001714.
  48. Kyowa Kirin GmbH (2020): Dossier zur Nutzenbewertung gemäß § 35a SGB V Burosumab (Crysvita) Modul 4A. [Zugriff: 09.11.2021]. URL: [https://www.g-ba.de/downloads/92-975-4126/2020-10-26\\_Modul4\\_Burosumab.pdf](https://www.g-ba.de/downloads/92-975-4126/2020-10-26_Modul4_Burosumab.pdf).
  49. Gemeinsamer Bundesausschuss (G-BA) (2021): Amendment zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo Wirkstoff: Burosumab. [Zugriff: 09.11.2021]. URL: [https://www.g-ba.de/downloads/92-975-4385/2021-04-15\\_Amendment-G-BA\\_%20Burosumab\\_D-588.pdf](https://www.g-ba.de/downloads/92-975-4385/2021-04-15_Amendment-G-BA_%20Burosumab_D-588.pdf).
  50. Ultragenyx Pharmaceutical Inc (2017): UX023-CL303 - Protocol and Protocol Amendments.

51. Ultragenyx Pharmaceutical Inc (2017): UX023-CL303 - Statistical Analysis Plan V 2.0.
52. Ultragenyx Pharmaceutical Inc (2015): Kit and Randomization Specification (KARS).
53. Skrinar A, Dvorak-Ewell M, Evins A, Macica C, Linglart A, Imel EA, et al. (2019): The lifelong impact of X-linked hypophosphatemia: results from a burden of disease survey. *Journal of the Endocrine Society*; 3(7):1321-34.
54. Nakamura Y, Takagi M, Takeda R, Miyai K, Hasegawa Y (2017): Hypertension is a characteristic complication of X-linked hypophosphatemia. *Endocrine journal*; 64(3):283-9.
55. Lecoq A-L, Trabado S, Schilbach K, Rothenbuhler A, Barosi A, Debza Y, et al. (2020): SUN-336 Obesity and Impaired Glucose Metabolism in Adult Patients with X-Linked Hypophosphatemia. *Journal of the Endocrine Society*; 4(Supplement\_1):SUN-336.
56. Lecoq A-L, Brandi ML, Linglart A, Kamenický P (2020): Management of X-linked hypophosphatemia in adults. *Metabolism*; 103:154049.
57. Seefried L, Smyth M, Keen R, Harvengt P (2021): Burden of disease associated with X-linked hypophosphatemia in adults: a systematic literature review. *Osteoporosis International*; 32:7-22.
58. Gemeinsamer Bundesausschuss (G-BA) (2020): Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 11 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 Verfo Wirkstoff: Burosumab. [Zugriff: 27.12.2021]. URL: [https://www.g-ba.de/downloads/92-975-3331/2019-10-01\\_Nutzenbewertung-G-BA\\_Burosumab\\_D-492.pdf](https://www.g-ba.de/downloads/92-975-3331/2019-10-01_Nutzenbewertung-G-BA_Burosumab_D-492.pdf).
59. Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG (2013): Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. *Cardiopulmonary physical therapy journal*; 24(3):21-9.
60. Bohannon RW, Crouch R (2016): Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. *Journal of evaluation in clinical practice*; 23(2):377-81.
61. Billeter R, Hoppeler H (2003): Muscular basis of strength. In: *Strength and power in sport*. 50-72.
62. Edman KP (2003): Contractile performance of skeletal muscle fibres. In: *Strength and power in sport*. 114-33.
63. Angst F, Aeschlimann A, Michel BA, Stucki G (2002): Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *The Journal of rheumatology*; 29(1):131-8.
64. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. (2008): Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *The journal of pain*; 9(2):105-21.
65. Javaid MK, Ward L, Pinedo-Villanueva R, Rylands AJ, Williams A, Insogna K, et al. (2021): Musculoskeletal features in adults with X-linked hypophosphatemia: An analysis of clinical trial and survey data. *The Journal of Clinical Endocrinology & Metabolism*;
66. Mindler GT, Kranzl A, Stauffer A, Kocijan R, Ganger R, Radler C, et al. (2021): Lower limb deformity and gait deviations among adolescents and adults with X-linked hypophosphatemia. *Frontiers in Endocrinology*:1170.

67. Kyowa Kirin Pharmaceutical Development Ltd (2021): A Pre-specified Analysis of the Maintenance of Effect of Burosumab Treatment Across all Time Points Collected from the 96-week Phase 3 UX023-CL303 Study through to Week 48 of the Phase 3b BUR02 Study in 31 Adults with X-linked Hypophosphataemia (XLH).
68. Ultragenyx Pharmaceutical Inc (2015): 2014-005529-11-GB Unique Product Identifier UX023-CL303 - A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study to Assess the Efficacy and Safety of KRN23 in Adults with X-linked Hypophosphatemia (XLH) - WHO ICTRP (EUCTR2014-005529-11-FR). Stand des Eintrags: 02.02.2019. [Zugriff: 02.11.2021]. URL: <https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2014-005529-11-FR>
69. Ultragenyx Pharmaceutical Inc (2015): 2014-005529-11-GB - A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study with Open-Label Extension to Assess the Efficacy and Safety of KRN23 in Adults with XLH- WHO ICTRP (EUCTR2014-005529-11-IE). Stand des Eintrags: 02.02.2019. [Zugriff: 02.11.2021]. URL: <https://trialssearch.who.int/Trial2.aspx?TrialID=EUCTR2014-005529-11-IE>

**Anhang 4-A: Suchstrategien – bibliografische Literaturrecherche**

Geben Sie nachfolgend die Suchstrategien für die bibliografische(n) Literaturrecherche(n) an, und zwar getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.). Für jede durchsuchte Datenbank ist die verwendete Strategie separat darzustellen. Geben Sie dabei zunächst jeweils den Namen der durchsuchten Datenbank (z. B. EMBASE), die verwendete Suchoberfläche (z. B. DIMDI, Ovid etc.), das Datum der Suche, das Zeitsegment (z. B.: „1980 to 2010 week 50“) und die gegebenenfalls verwendeten Suchfilter (mit Angabe einer Quelle) an. Listen Sie danach die Suchstrategie einschließlich der resultierenden Trefferzahlen auf. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel (eine umfassende Suche soll Freitextbegriffe und Schlagwörter enthalten):

<b>Datenbankname</b>	EMBASE	
<b>Suchoberfläche</b>	Ovid	
<b>Datum der Suche</b>	07.11.2016	
<b>Zeitsegment</b>	1974 to 2016 November 04	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle <sup>23</sup> ] – Strategy minimizing difference between sensitivity and specificity	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	Diabetes Mellitus/	552986
2	Non Insulin Dependent Diabetes Mellitus/	195234
3	(diabet* or niddm or t2dm).ab,ti.	714228
4	or/1-3	847068
5	linagliptin*.mp.	1562
6	(random* or double-blind*).tw.	1193849
7	placebo*.mp.	388057
8	or/6-7	1382838
9	and/4,5,8	633

<sup>23</sup> Das Zitat zu dem hier beispielhaft angegebenen Suchfilter lautet wie folgt: Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. J Med Libr Assoc 2006; 94(4): 451-455. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

**Anhang 4-A1: Suche nach RCT mit dem zu bewertenden Arzneimittel**

Tabelle 4-84: Dokumentation der Suchstrategie für randomisierte kontrollierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in EMBASE

<b>Datenbankname</b>	EMBASE	
<b>Suchoberfläche</b>	Ovid	
<b>Datum der Suche</b>	02.11.2021	
<b>Zeitsegment</b>	1974 to 2021 November 01	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 – Strategy minimizing difference between sensitivity and specificity	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
<b>1</b>	exp burosumab/	257
<b>2</b>	krn23.mp. <sup>a</sup>	44
<b>3</b>	"krn 23".mp. <sup>a</sup>	19
<b>4</b>	burosumab.mp.	275
<b>5</b>	anti-FGF23.mp.	122
<b>6</b>	"anti-FGF 23".mp.	18
<b>7</b>	crysvita.mp.	28
<b>8</b>	1 or 2 or 3 or 4 or 5 or 6 or 7	354
<b>9</b>	random*.tw. or placebo*.mp. or double-blind*.tw.	1984382
<b>10</b>	8 und 9	75
a: für krn 23 bzw. krn23 lag kein MeSH-Term vor.		

Tabelle 4-85: Dokumentation der Suchstrategie für randomisierte kontrollierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in MEDLINE

<b>Datenbankname</b>	MEDLINE	
<b>Suchoberfläche</b>	Ovid	
<b>Datum der Suche</b>	02.11.2021	
<b>Zeitsegment</b>	Ovid MEDLINE(R) and In-Process, In-Data-Review & Other Non-Indexed Citations and Daily 1946 to November 01, 2021	
<b>Suchfilter</b>	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle <sup>24</sup> ] – Strategy minimizing difference between sensitivity and specificity	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	burosumab.mp.	103
2	krn23.mp.	16
3	"krn 23".mp.	1
4	anti-FGF23.mp.	57
5	"anti-FGF 23".mp.	8
6	crysvita.mp.	6
7	1 or 2 or 3 or 4 or 5 or 6	142
8	randomized controlled trial.pt. or randomized.mp. or placebo.mp.	979763
9	7 and 8	11

<sup>24</sup> Das Zitat zu dem hier beispielhaft angegebenen Suchfilter lautet wie folgt: Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. J Med Libr Assoc 2006; 94(4): 451-455. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

Tabelle 4-86: Dokumentation der Suchstrategie für randomisierte kontrollierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in Cochrane

<b>Datenbankname</b>	Cochrane	
<b>Suchoberfläche</b>	Cochrane Library	
<b>Datum der Suche</b>	02.11.2021	
<b>Zeitsegment</b>	Keine Einschränkung	
<b>Suchfilter</b>	Keine Suchfilter	
<b>#</b>	<b>Suchbegriffe</b>	<b>Ergebnis</b>
1	krn23	21
2	"krn 23"	2
3	anti-FGF23	27
4	"anti-FGF 23"	2
5	crysvita	0
6	burosumab	54
7	#1 or #2 or #3 or #4 or #5 or #6	69

Am 02.11.2021 fand eine bibliographische Literaturrecherche für RCT mit Burosumab statt. Es wurden insgesamt 155 Treffer identifiziert. Hiervon wurden 80 Treffer ausgeschlossen, da es sich hierbei um Dubletten handelt. Von den verbliebenen 75 Publikationen wurden 72 Publikationen mittels Titel- und Abstractscreening auf Basis der prädefinierten Ein- und Ausschlusskriterien ausgeschlossen (Tabelle 4-5 und Tabelle 4-6). Insgesamt wurden somit drei relevante Treffer identifiziert. Dabei handelt es sich um die Publikationen zu der in diesem Dossier dargestellten pivotalen Studie UX023-CL303 [4, 23, 47].

#### **Anhang 4-A2: Suche nach RCT für indirekte Vergleiche**

Entfällt. Es werden keine indirekten Vergleiche auf Basis von RCTs durchgeführt.

#### **Anhang 4-A3: Suche nach nicht randomisierten vergleichenden Studien**

Entfällt. Es werden keine weiteren nicht randomisierten vergleichenden Studien im vorliegenden Dossier dargestellt.

#### **Anhang 4-A4: Suche nach weiteren Untersuchungen**

Entfällt. Es werden keine weiteren Untersuchungen im vorliegenden Dossier durchgeführt.

**Anhang 4-B: Suchstrategien – Suche in Studienregistern/ Studienergebnisdatenbanken**

Geben Sie nachfolgend die Suchstrategien für die Suche(n) in Studienregistern/ Studienergebnisdatenbanken an. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Für jede/s durchsuchte Studienregister/ Studienergebnisdatenbank ist eine separate Strategie darzustellen. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters/ Studienergebnisdatenbank (z. B. clinicaltrials.gov), die Internetadresse, unter der das/die Studienregister/ Studienergebnisdatenbank erreichbar ist (z. B. <http://www.clinicaltrials.gov>), das Datum der Suche, die verwendete Suchstrategie und die resultierenden Treffer an. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

<b>Studienregister/ Studienergebnisdatenbank</b>	International Clinical Trials Registry Platform Search Portal
<b>Internetadresse</b>	<a href="http://apps.who.int/trialsearch/">http://apps.who.int/trialsearch/</a>
<b>Datum der Suche</b>	07.11.2016
<b>Eingabeoberfläche</b>	Standard Search
<b>Suchstrategie</b>	linagliptin OR BI 1356
<b>Treffer</b>	169

**Anhang 4-B1: Suche nach RCT mit dem zu bewertenden Arzneimittel**

Tabelle 4-87: Dokumentation der Suchstrategie für randomisierte Studien mit Burosumab (Crysvita®) zur Behandlung der XLH in clinicaltrials.gov

<b>Studienregister/ Studienergebnisdatenbank</b>	clinicaltrials.gov
<b>Internetadresse</b>	<a href="https://clinicaltrials.gov/">https://clinicaltrials.gov/</a>
<b>Datum der Suche</b>	02.11.2021
<b>Suchstrategie</b>	[OTHER TERMS] KRN23 OR "KRN 23" OR burosumab OR "anti-FGF 23" OR anti-FGF23 OR crysvita
<b>Treffer</b>	31

Die Suche in clinicaltrials.gov ergab 31 Treffer. Unter den Treffern befindet sich auch die relevante pivotale Studie UX023-CL303, die in Abschnitt 4.3.1.3 dargestellt ist [10]. Die übrigen Treffer erfüllten nicht die prädefinierten Einschlusskriterien (Tabelle 4-5).

Tabelle 4-88: Dokumentation der Suchstrategie für randomisierte Studien mit Burosumab (Crysvita<sup>®</sup>) zur Behandlung der XLH in EU Clinical Trial Register (EU-CTR)

<b>Studienregister/ Studienergebnisdatenbank</b>	EU Clinical Trials Register (EU-CTR)
<b>Internetadresse</b>	<a href="https://www.clinicaltrialsregister.eu/ctr-search/search">https://www.clinicaltrialsregister.eu/ctr-search/search</a>
<b>Datum der Suche</b>	02.11.2021
<b>Suchstrategie</b>	KRN23 OR "KRN 23" OR burosumab OR "anti-FGF 23" OR anti-FGF23 OR crysvita
<b>Treffer</b>	8

Die Suche in *EU Clinical Trials Register* ergab 8 Treffer. Unter den Treffern befindet sich auch die relevante pivotale Studie UX023-CL303, die in Abschnitt 4.3.1.3 dargestellt ist [11]. Die übrigen Treffer erfüllten nicht die prädefinierten Einschlusskriterien (Tabelle 4-5).

Tabelle 4-89: Dokumentation der Suchstrategie für randomisierte Studien mit Burosumab zur Behandlung der XLH (Crysvita<sup>®</sup>) in *WHO International Clinical Trials Registry Platform* (WHO ICTRP)

<b>Studienregister/ Studienergebnisdatenbank</b>	WHO International Clinical Trials Registry Platform (WHO ICTRP)
<b>Internetadresse</b>	<a href="https://trialsearch.who.int/Default.aspx">https://trialsearch.who.int/Default.aspx</a>
<b>Datum der Suche</b>	02.11.2021
<b>Suchstrategie</b>	KRN23 OR burosumab OR anti-FGF 23 OR anti-FGF23 OR crysvita
<b>Treffer</b>	36

Die Suche im *International Clinical Trials Registry Platform Search Portal* der WHO ergab 58 Einträge für 36 Studien. Unter den Treffern befindet sich auch die relevante pivotale Studie UX023-CL303, die in Abschnitt 4.3.1.3 dargestellt ist. Für diese Studie existieren insgesamt 3 Studienregistereinträge [12, 68, 69]. Dabei handelt es sich jedoch um Mehrfacheinträge der gleichen Studie. Die übrigen Treffer erfüllten nicht die prädefinierten Einschlusskriterien (Tabelle 4-5).

#### **Anhang 4-B2: Suche nach RCT für indirekte Vergleiche**

Entfällt. Es wird kein indirekter Vergleich auf Basis randomisierter kontrollierter Studien durchgeführt.

#### **Anhang 4-B3: Suche nach nicht randomisierten vergleichenden Studien**

Entfällt. Es liegen keine nicht randomisierten vergleichenden Studien in dem hier betrachteten Anwendungsgebiet von Burosumab vor.

#### **Anhang 4-B4: Suche nach weiteren Untersuchungen**

Entfällt. Es wurden keine weiteren Untersuchungen in dem hier betrachteten Anwendungsgebiet von Burosumab durchgeführt.

**Anhang 4-C: Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche)**

*Listen Sie nachfolgend die im Volltext gesichteten und ausgeschlossenen Dokumente aus der /den bibliografischen Literaturrecherche(n) auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.*

**Anhang 4-C1: Suche nach RCT mit dem zu bewertenden Arzneimittel**

Die Ausschlusskriterien sind in Tabelle 4-6 aufgeführt.

Im Rahmen der bibliographischen Literaturrecherche wurden keine Dokumente im Volltext gesichtet und ausgeschlossen.

**Anhang 4-C2: Suche nach RCT für indirekte Vergleiche**

Entfällt. Es werden keine indirekten Vergleiche auf Basis von RCTs durchgeführt.

**Anhang 4-C3: Suche nach nicht randomisierten vergleichenden Studien**

Entfällt. Es werden keine weiteren nicht randomisierten vergleichenden Studien im vorliegenden Dossier dargestellt.

**Anhang 4-C4: Suche nach weiteren Untersuchungen**

Entfällt. Es werden keine weiteren Untersuchungen im vorliegenden Dossier dargestellt.

#### Anhang 4-D: Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern/ Studienergebnisdatenbanken)

Listen Sie nachfolgend die durch die Studienregistersuche(n)/ Studienergebnisdatenbanksuche(n) identifizierten, aber ausgeschlossenen Registereinträge auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

#### Anhang 4-D1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Tabelle 4-90: Zusammenfassende Darstellung der aus- und eingeschlossenen Studien

Studienregister	Einträge	Ausgeschlossen	Eingeschlossen
clinicaltrials.gov	31	30	1
EU-CTR	8	7	1
ICTRP	36	33	3 <sup>a</sup>
<b>Gesamt</b>	<b>75</b>	<b>70</b>	<b>5<sup>a</sup></b>

a: Hierbei handelt es sich jeweils um die Studie UX023-CL303, die in den verschiedenen Registern aufgeführt ist und bei WHO ICTRP insgesamt drei Mal aufgeführt ist.

Tabelle 4-91: Ausgeschlossene Studien ClinicalTrials.gov

Nr.	Titel	Register-nummer	Ausschlussgrund
1	An Open-label, Multi-center, Single-cohort, Post-marketing Phase 4 Study to Evaluate the Efficacy, Pharmacodynamics, and Safety of the Anti-FGF23 Antibody, KRN23, in Pediatric Chinese Patients With X-linked Hypophosphatemic Rickets/Osteomalacia (XLH) - ClinicalTrials.gov	NCT04842032	A5
2	An Open-label, Multi-center, Single-cohort, Post-marketing Phase 4 Study to Evaluate the Efficacy, Pharmacodynamics, and Safety of the Anti-FGF23 Antibody, KRN23, in Adult Chinese Patients With X-linked Hypophosphatemic Rickets/Osteomalacia - ClinicalTrials.gov	NCT04842019	A5
3	An Investigator-sponsored Phase 3b Open-label Study of Anti-FGF23 Antibody Burosumab (KRN23) in Adult Patients With Xlinked Hypophosphatemia (XLH) in GERmany - BurGER - ClinicalTrials.gov	NCT04695860	A5
4	A Phase 3 Long-term Extension Study of KRN23 in Patients With X-linked Hypophosphatemic Rickets/Osteomalacia - ClinicalTrials.gov	NCT04308096	A5
5	An Open Label Trial to Assess the Safety and Efficacy of KRN23, an Investigational Antibody to FGF23, in a Single Pediatric Patient With Epidermal Nevus Syndrome (ENS) and Associated Hypophosphatemic Rickets - ClinicalTrials.gov	NCT04320316	A5

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Nr.	Titel	Register- nummer	Ausschluss- grund
6	Examining the Effect of Burosumab on Muscle Function Using MR Spectroscopy - ClinicalTrials.gov	NCT04146935	A5
7	A Phase 1/2, Open-label, Multicenter, Non-randomized Study to Assess the Safety, Tolerability, Pharmacokinetics and Efficacy of Burosumab in Pediatric Patients From Birth to Less Than 1 Year of Age With X-linked Hypophosphatemia (XLH) - ClinicalTrials.gov	NCT04188964	A5
8	An Open Label Trial to Assess the Safety and Efficacy of Burosumab in a Single Patient With Cutaneous Skeletal Hypophosphatemia Syndrome (CSHS) - ClinicalTrials.gov	NCT03993821	A5
9	A Phase 3b Open-label Study of the Anti-FGF23 Antibody, Burosumab (KRN23) in Adult Patients With X-linked Hypophosphatemia (XLH) - ClinicalTrials.gov	NCT03920072	A5
10	12-months of Treatment With Burosumab in Children and Adolescents With X-linked Hypophosphatemia: a Prospective Longitudinal Cohort Study - ClinicalTrials.gov	NCT04419363	A1
11	An Open Label Trial to Assess the Safety and Efficacy of Burosumab (KRN23), an Investigational Antibody to FGF23, in a Single Pediatric Patient With Epidermal Nevus Syndrome(ENS) and Associated Hypophosphatemic Rickets - ClinicalTrials.gov	NCT03581591	A5
12	A Phase 3 Open-Label Trial to Assess the Efficacy and Safety of KRN23 in Pediatric Patients With X-linked Hypophosphatemic Rickets/Osteomalacia - ClinicalTrials.gov	NCT03233126	A5
13	A Randomized, Open-Label, Phase 3 Study to Assess the Efficacy and Safety of KRN23 Versus Oral Phosphate and Active Vitamin D Treatment in Pediatric Patients With X Linked Hypophosphatemia (XLH) - ClinicalTrials.gov	NCT02915705	A5
14	A Phase 2 Open-Label Trial to Assess the Efficacy and Safety of KRN23 in Patients With Tumor-Induced Osteomalacia or Epidermal Nevus Syndrome - ClinicalTrials.gov	NCT02722798	A5
15	An Open-Label, Phase 2 Study to Assess the Safety, Pharmacodynamics, and Efficacy of KRN23 in Children From 1 to 4 Years Old With X-linked Hypophosphatemia (XLH) - ClinicalTrials.gov	NCT02750618	A5
16	An Open-Label, Single-Arm, Phase 3 Study to Evaluate the Effects of KRN23 on Osteomalacia in Adults With X-linked Hypophosphatemia (XLH) - ClinicalTrials.gov	NCT02537431	A5
17	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) - ClinicalTrials.gov	NCT02526160	E
18	A Phase 2b, Open-Label, Long-Term Extension Study to Evaluate the Safety and Pharmacodynamics of KRN23 in Adult Subjects With X-Linked Hypophosphatemia (XLH) - ClinicalTrials.gov	NCT02312687	A5
19	A Phase 2 Open-Label Trial to Assess the Efficacy and Safety of KRN23, an Antibody to FGF23, in Subjects With Tumor-Induced Osteomalacia (TIO) or Epidermal Nevus Syndrome (ENS)-Associated Osteomalacia - ClinicalTrials.gov	NCT02304367	A5

---

 Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen
 

---

Nr.	Titel	Register-nummer	Ausschluss-grund
20	A Randomized, Open-Label, Dose Finding, Phase 2 Study to Assess the Pharmacodynamics and Safety of the Anti-FGF23 Antibody, KRN23, in Pediatric Patients With X-linked Hypophosphatemia (XLH) - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT02163577	A5
21	A Phase 1, Multicenter, Open-label, Sequential Dose-escalation, Single-dose Study to Assess the Safety and Tolerability of KRN23 in Subjects With X-linked Hypophosphatemic Rickets/Osteomalacia. - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT02181764	A5
22	An Open-Label, Long-Term, Extension Study to Evaluate the Safety and Efficacy of KRN23 in Adult Subjects With X-Linked Hypophosphatemia - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT01571596	A5
23	A Phase I/II, Open-Label, Repeat-Dose, Dose-Escalation Study of KRN23 in Adult Subjects With X-Linked Hypophosphatemia - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT01340482	A5
24	Expanded Access to Burosumab	NCT03775187	A5
25	Effect of Burosumab and 1-25 (OH) Vitamin D on Human Osteoblasts From Patients Requiring Craniosynostosis Surgery for Idiopathic Reason or Due to Hypophosphatemic Rickets (HR) - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT04159675	A5
26	A Phase I, Double-blind, Randomized, Placebo-controlled, Single-dose, Dose-escalation Study of KRN23 in X-linked Hypophosphatemia - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT00830674	A6
27	Effect of Burosumab and 1-25 (OH) Vitamin D on Human Osteoclasts From Patients With Hypophosphatemic Rickets (HR) - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT04184661	A5
28	X-linked Hypophosphatemia Disease Monitoring Program (XLH-DMP) - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT03651505	A3
29	Tumor-induced Osteomalacia Disease Monitoring Program (TIO DMP) - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT04783428	A1
30	A UK Multicentre, Non-interventional, Observational, Health-Related Quality of Life Study for Children and Adolescents With X-linked Hypophosphataemia - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT04819490	A5
31	An International, Multicentre, Prospective, Non-interventional Observational Registry for Patients With X-Linked Hypophosphatemia (XLH) - <a href="https://ClinicalTrials.gov">ClinicalTrials.gov</a>	NCT03193476	A5

Tabelle 4-92: Ausgeschlossene Studien EU-CTR

Nr.	Titel	Register-nummer	Ausschluss-grund
32	Hypophosphatemia (XLH) - EU-CTR	2018-000202-37	A5
33	An investigator-sponsored Phase 3b Open-label Study of Anti-FGF23 Antibody Burosumab (KRN23) in Adult Patients with X-linked Hypophosphatemia (XLH) in GERmany - BurGER - EU-CTR	2019-003190-26	A5
34	A Phase 1/2, Open-label, Multicenter, Non-randomized Study to Assess the Safety, Tolerability, Pharmacokinetics and Efficacy of Burosumab in Pediatric Patients from Birth to Less than 1 Year of Age with X-linked Hypophosphatemia (XLH) - EU-CTR	2019-000469-19	A5
35	A Randomized, open Label, Dose Finding, Phase 2 Study to Assess the Pharmacodynamics and Safety of the anti-FGF23 antibody, KRN23, in Pediatric Patients with X-linked Hypophosphatemia (XLH) - EU-CTR	2014-000406-35	A5
36	An Open-Label, Phase 2 Study to Assess the Safety, Pharmacodynamics, and Efficacy of KRN23 in Children from 1 to 4 Years Old with X-linked Hypophosphatemia (XLH) - EU-CTR	2018-001983-49	A5
37	A Randomized, Open-Label, Phase 3 Study to Assess the Efficacy and Safety of KRN23 Versus Oral Phosphate and Active Vitamin D Treatment in Pediatric Patients with X-linked Hypophosphatemia (XLH)	2016-000600-29	A3
38	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) - EU-CTR	2014-005529-11	E
39	An Open-Label, Single-Arm, Phase 3 Study to Evaluate the Effects of KRN23 on Osteomalacia in Adults with X-linked Hypophosphatemia (XLH) - EU-CTR	2015-001775-41	A5

Tabelle 4-93: Ausgeschlossene Studien ICTRP

Nr.	Titel	Register-nummer	Ausschluss-grund
40	An Open-label, Multi-center, Single-cohort, Post-marketing Phase 4 Study to Evaluate the Efficacy, Pharmacodynamics, and Safety of the Anti-FGF23 Antibody, KRN23, in Adult Chinese Patients With X-linked Hypophosphatemic Rickets/Osteomalacia - WHO ICTRP	NCT04842019	A5
41	An Open-label, Multi-center, Single-cohort, Post-marketing Phase 4 Study to Evaluate the Efficacy, Pharmacodynamics, and Safety of the Anti-FGF23 Antibody, KRN23, in Pediatric Chinese Patients With X-linked Hypophosphatemic Rickets/Osteomalacia (XLH) - WHO ICTRP	NCT04842032	A5
42	An Investigator-sponsored Phase 3b Open-label Study of Anti-FGF23 Antibody Burosumab (KRN23) in Adult Patients With Xlinked Hypophosphatemia (XLH) in GERmany - BurGER - WHO ICTRP	NCT04695860	A5
43	An investigator-sponsored Phase 3b Open-label Study of Anti-FGF23 Antibody Burosumab (KRN23) in Adult Patients with X-linked Hypophosphatemia (XLH) in GERmany - BurGER - WHO	2019-003190-26	A5

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Nr.	Titel	Register- nummer	Ausschluss- grund
	ICTRP		
44	12-months of Treatment With Burosumab in Children and Adolescents With X-linked Hypophosphatemia: a Prospective Longitudinal Cohort Study - WHO ICTRP	NCT04419363	A1
45	Phase 3 Long-term Extension Study of KRN23 in Patients with X-linked Hypophosphatemic Rickets/Osteomalacia - WHO ICTRP	JPRN- JapicCTI- 205284	A1
46	An Open Label Trial to Assess the Safety and Efficacy of KRN23, an Investigational Antibody to FGF23, in a Single Pediatric Patient With Epidermal Nevus Syndrome (ENS) and Associated Hypophosphatemic Rickets - WHO ICTRP	NCT04320316	A5
47	A Phase 3 Long-term Extension Study of KRN23 in Patients With X-linked Hypophosphatemic Rickets/Osteomalacia - WHO ICTRP	NCT04308096	A5
48	Effect of Burosumab and 1-25 (OH) Vitamin D on Human Osteoclasts From Patients With Hypophosphatemic Rickets (HR) - WHO ICTRP	NCT04184661	A5
49	A Phase 1/2, Open-label, Multicenter, Non-randomized Study to Assess the Safety, Tolerability, Pharmacokinetics and Efficacy of Burosumab in Pediatric Patients From Birth to Less Than 1 Year of Age With X-linked Hypophosphatemia (XLH) - WHO ICTRP	NCT04188964	A5
50	Effect of Burosumab and 1-25 (OH) Vitamin D on Human Osteoblasts From Patients Requiring Craniostylosis Surgery for Idiopathic Reason or Due to Hypophosphatemic Rickets (HR) - WHO ICTRP	NCT04159675	A5
51	A Phase 1/2, Open-label, Multicenter, Non-randomized Study to Assess the Safety, Tolerability, Pharmacokinetics and Efficacy of Burosumab in Pediatric Patients from Birth to Less than 1 Year of Age with X-linked Hypophosphatemia (XLH) - WHO ICTRP	2019-000469- 19	A5
52	Examining the Effect of Burosumab on Muscle Function Using MR Spectroscopy - WHO ICTRP	NCT04146935	A5
53	A Phase 3b Open-label Study of the Anti-FGF23 Antibody, - WHO ICTRP	2018-000202- 37	A5
54	An Open Label Trial to Assess the Safety and Efficacy of Burosumab in a Single Patient With Cutaneous Skeletal Hypophosphatemia Syndrome (CSHS) - WHO ICTRP	NCT03993821	A5
55	Expanded Access to Burosumab - WHO ICTRP	NCT03775187	A5
56	A Phase 3b Open-label Study of the Anti-FGF23 Antibody, Burosumab (KRN23) in Adult Patients With X-linked Hypophosphatemia (XLH) - WHO ICTRP	NCT03920072	A5
57	A Phase 3b Open-label Study of the Anti-FGF23 Antibody, Burosumab (KRN23) in Adult Patients with X-linked Hypophosphatemia (XLH) - WHO ICTRP	2018-000202- 37	A5
58	An Open Label Trial to Assess the Safety and Efficacy of Burosumab (KRN23), an Investigational Antibody to FGF23, in a Single Pediatric Patient With Epidermal Nevus Syndrome (ENS) and Associated Hypophosphatemic Rickets - WHO ICTRP	NCT03581591	A5

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Nr.	Titel	Register- nummer	Ausschluss- grund
59	A Phase 3 Open-Label Trial to Assess the Efficacy and Safety of KRN23 in Pediatric Patients with X-linked Hypophosphatemic Rickets/Osteomalacia - WHO ICTRP	JPRN- JapicCTI- 173614	A5
60	A Randomized, Open-Label, Phase 3 Study to Assess the Efficacy and Safety of KRN23 - WHO ICTRP	2016-000600- 29	A3
61	A Phase 2 Open-Label Trial to Assess the Efficacy and Safety of KRN23 in Patients With Tumor-Induced Osteomalacia or Epidermal Nevus Syndrome - WHO ICTRP	NCT02722798	A5
62	An Open-Label, Phase 2 Study to Assess the Safety, Pharmacodynamics, and Efficacy of KRN23 in Children From 1 to 4 Years Old With X-linked Hypophosphatemia (XLH) - WHO ICTRP	NCT02750618	A5
63	A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study to Assess the Efficacy and Safety of KRN23 in Adults with X-linked Hypophosphatemia (XLH) - WHO ICTRP	2014-005529- 11	E
64	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) - WHO ICTRP	2014-005529- 11	E
65	An Open-Label, Single-Arm, Phase 3 Study to Evaluate the Effects of KRN23 on Osteomalacia in Adults With X-linked Hypophosphatemia (XLH) - WHO ICTRP	NCT02537431	A5
66	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) - WHO ICTRP	NCT02526160	E
67	A Randomized, open Label, Dose Finding, Phase 2 Study to Assess the Pharmacodynamics and Safety of the anti-FGF23 antibody, KRN23, in Pediatric Patients with X-linked Hypophosphatemia (XLH) - WHO ICTRP	2014-000406- 35	A5
68	A Phase 2b, Open-Label, Long-Term Extension Study to Evaluate the Safety and Pharmacodynamics of KRN23 in Adult Subjects With X-Linked Hypophosphatemia (XLH) - WHO ICTRP	NCT02312687	A5
69	A Phase 2 Open-Label Trial to Assess the Efficacy and Safety of KRN23, an Antibody to FGF23, in Subjects With Tumor-Induced Osteomalacia (TIO) or Epidermal Nevus Syndrome (ENS)-Associated Osteomalacia - WHO ICTRP	NCT02304367	A5
70	A Phase 1, Multicenter, Open-label, Sequential Dose-escalation, Single-dose Study to Assess the Safety and Tolerability of KRN23 in Subjects With X-linked Hypophosphatemic Rickets/Osteomalacia. - WHO ICTRP	NCT02181764	A5
71	A Phase 1 study of KRN23 in Subjects with X-linked Hypophosphatemic Rickets/Osteomalacia - WHO ICTRP	JPRN- JapicCTI- 142601	A5
72	A Randomized, open Label, Dose Finding, Phase 2 Study to Assess the Pharmacodynamics and Safety of the anti-FGF23 antibody, KRN23, in Pediatric Patients with X-linked Hypophosphatemia (XLH) - WHO ICTRP	2014-000406- 35	A5
73	An Open-Label, Long-Term, Extension Study to Evaluate the Safety and Efficacy of KRN23 in Adult Subjects With X-Linked Hypophosphatemia - WHO ICTRP	NCT01571596	A5

Nr.	Titel	Register- nummer	Ausschluss- grund
74	A Phase I/II, Open-Label, Repeat-Dose, Dose-Escalation Study of KRN23 in Adult Subjects With X-Linked Hypophosphatemia - WHO ICTRP	NCT01340482	A5
75	A Phase I, Double-blind, Randomized, Placebo-controlled, Single-dose, Dose-escalation Study of KRN23 in X-linked Hypophosphatemia - WHO ICTRP	NCT00830674	A6

**Anhang 4-D2: Suche nach RCT für indirekte Vergleiche**

Entfällt. Es werden keine indirekten Vergleiche durchgeführt.

**Anhang 4-D3: Suche nach nicht randomisierten vergleichenden Studien**

Entfällt. Es werden keine weiteren nicht randomisierten vergleichenden Studien im vorliegenden Dossier dargestellt.

**Anhang 4-D4: Suche nach weiteren Untersuchungen**

Entfällt. Es werden keine weiteren Untersuchungen im vorliegenden Dossier dargestellt.

**Anhang 4-E: Methodik der eingeschlossenen Studien – RCT**

*Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.5 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-94 inklusive eines Flow-Charts für den Patientenfluss.*

*Sollten Sie im Dossier indirekte Vergleiche präsentieren, beschreiben Sie ebenfalls die Methodik jeder zusätzlich in den indirekten Vergleich eingeschlossenen Studie (Abschnitt 4.3.2.1). Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-94 inklusive eines Flow-Charts für den Patientenfluss.*

Tabelle 4-94 (Anhang): Studiendesign und -methodik für Studie UX023-CL303

Item <sup>a</sup>	Charakteristikum	Studieninformation
<b>Studienziel</b>		
<b>2 b</b>	Genaue Ziele, Fragestellung und Hypothesen	Ziel dieser randomisierten, Placebo-kontrollierten doppelblinden multizentrischen Studie der Phase III war die Untersuchung der Wirksamkeit und Sicherheit von Burosumab bei erwachsenen XLH-Patienten (18 – 65 Jahre) im Vergleich zu einer Placebo-Kontrolle. Hierbei sollte spezifisch das Ausmaß des Zusatznutzens von Burosumab gegenüber Placebo anhand von patientenrelevanten Endpunkten evaluiert werden. Die Hypothese war, dass die Behandlung mit 1 mg/kg Körpergewicht (KG) Burosumab effektiver als Placebo hinsichtlich des Erreichens einer <i>Serumphosphat-Konzentration</i> (mg/dl) in der Mitte des Dosisintervalls oberhalb der unteren Grenze des Normbereichs von 2,5 mg/dl ist, gemittelt über die Dosiszyklen zwischen Baseline und Woche 24.
<b>Methoden</b>		
<b>3</b>	Studiendesign	
<b>3a</b>	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Es handelt sich um eine multizentrische, randomisierte, doppelblinde Placebo-kontrollierte Studie der Phase III mit einem Zuteilungsverhältnis von 1:1 und einem 4-wöchigen Applikationsintervall von Burosumab bzw. Placebo.  Die Studie beinhaltet 2 Screeningvisiten sowie eine Baseline-Visite mit einer anschließenden 24-wöchigen Placebo-kontrollierten doppelblinden Behandlungsphase. An diese doppelblinde Behandlungsphase schloss sich eine 24-wöchige offene Behandlungsphase (Behandlungsfortsetzungsphase) an, in der die Studienteilnehmer aus der Placebo-Gruppe in die Burosumab-Gruppe wechselten, sodass alle Studienteilnehmer ab diesem Zeitpunkt Burosumab erhielten. Die ursprüngliche Behandlungszuteilung wurde dem Patienten oder dem Prüfarzt nicht offenbart. Anschließend fand eine 48-wöchige Extensionsphase I statt, um die Effizienz und Sicherheit der Langzeitbehandlung zu evaluieren. Zusätzlich hatten die Studienteilnehmer der USA die Gelegenheit, an einer weiteren Extensionsphase

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>(Extensionsphase II) für bis zu 53 Wochen teilzunehmen.</p> <p><u>Behandlungsphase Burosumab</u> <u>Woche 0 – 24:</u> Die Initialdosis von Burosumab betrug 1,0 mg/kg Körpergewicht s.c. alle vier Wochen.</p> <p><u>Behandlungsphase Placebo Woche 0 – 24:</u> Placebo wurde ebenfalls wie Burosumab s.c. alle vier Wochen verabreicht.</p> <p><u>Behandlungsfortsetzungsphase</u> <u>Woche 25 – 48:</u> Alle Studienteilnehmer erhielten Burosumab s.c. alle vier Wochen.</p> <p><u>Extensionsphase I Woche 49 – 96:</u> Alle Studienteilnehmer erhielten Burosumab s.c. alle vier Wochen.</p> <p><u>Extensionsphase II Woche 97 – 149:</u> Alle Studienteilnehmer (USA) erhielten Burosumab s.c. alle vier Wochen.</p>
<b>3b</b>	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<p>Die Randomisierung der Studienteilnehmer erfolgte 1:1 in die beiden Behandlungsgruppen (Burosumab und Placebo) und wurde anhand des <i>Interactive Web Response System (IWRS)</i> durchgeführt [52]. Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 <i>Brief Pain Inventory (BPI) Schlimmster Schmerz</i> des patientenberichteten Fragebogens <i>BPI</i> und der Region erfolgen. Der Stratifizierungsfaktor Region wurde aufgrund von operativen und logistischen Gründen zusätzlich zu dem Merkmal Schmerz mit aufgenommen. Da lediglich eine geringe Anzahl an Studienteilnehmern aus Japan und Korea zu erwarten war, sollte dies eine Gleichverteilung auf die beiden Behandlungsgruppen Burosumab und Placebo gewährleisten. Der Stratifizierungsfaktor Schmerz wurde anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens <i>BPI</i> bestimmt. Die verkürzte Form des Fragebogens enthielt die Fragen 3 – 6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>der einzelnen Items, sodass Frage 3 (<i>BPI Schlimmster Schmerz</i>) als Frage 1 klassifiziert wurde und Frage 5 (<i>BPI Durchschnittlicher Schmerz</i>) als Frage 3. Das für die Stratifizierung verwendete <i>Interactive web response system (IWRS)</i> war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (<i>BPI Schlimmster Schmerz</i>) sondern Frage 3 (<i>BPI Durchschnittlicher Schmerz</i>) verwendet wurde (Anhang 4-E 8b).</p> <p>Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen, mit der Stratifizierung fortzufahren. Dies wurde als Protokollabweichung im Electronic Data Capture System (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien <i>BPI Durchschnittlicher Schmerz</i> <math>\leq 6,0</math> bzw. <i>BPI Durchschnittlicher Schmerz</i> <math>&gt; 6,0</math> eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien <i>BPI Schlimmster Schmerz</i> <math>\leq 6,0</math> und <i>BPI Schlimmster Schmerz</i> <math>&gt; 6,0</math>. Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (<i>BPI Durchschnittlicher Schmerz</i> <math>\leq 6,0</math> und <i>BPI Durchschnittlicher Schmerz</i> <math>&gt; 6,0</math>) dem schlimmsten Schmerz (<i>BPI Schlimmster Schmerz</i> <math>\leq 6,0</math> und <i>BPI Schlimmster Schmerz</i> <math>&gt; 6,0</math>). Der Pearson Korrelationskoeffizient <math>r</math> zwischen den Gruppen betrug 0,82256 (<math>p &lt; 0,001</math>) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Des Weiteren erfolgten Sensitivitätsanalysen des primären und der drei sekundären Wirksamkeitsendpunkte um die Robustheit der Ergebnisse zu überprüfen, die ebenfalls bestätigten, dass der Fehler in der Stratifizierung zu keiner Verzerrung der Ergebnisse führte.</p> <p><u>Amendment 1 datiert auf den 21.06.2015:</u> <u>Studiendesign:</u></p> <ul style="list-style-type: none"> <li>Die 24-wöchige offene Behandlungsfortsetzungsphase wurde um eine 48-wöchige Extensionsphase erweitert. Dementsprechend wurde der Titel modifiziert.</li> </ul>

		<p><u>Änderung der Ein- und Ausschlusskriterien:</u></p> <ul style="list-style-type: none"> <li>• Das Einschlusskriterium #5 wurde abgeändert: Die geschätzte glomeruläre Filtrationsrate (eGFR) sollte bei <math>\geq 60</math> ml/min liegen (<i>Chronic kidney disease epidemiology collaboration</i> (CDK-EPI)); bei einem eGFR von 45 bis <math>&lt; 60</math> mL/min am Tag der Screeningvisite 2 sollte eine Bestätigung vorliegen, dass die Niereninsuffizienz nicht durch eine Nephrokalzinose verursacht wurde.</li> <li>• Das Einschlusskriterium #10 wurde erweitert. Studienteilnehmer die sich nicht einer vollständigen Hysterektomie unterzogen haben, müssen 2 effektive Verhütungsmethoden anwenden.</li> <li>• Das Ausschlusskriterium #8 wurde dahingehend modifiziert, dass nicht nur XLH-Patienten die oral verabreichte Biphosphonate sondern jegliche Biphosphonate innerhalb der letzten 2 Jahre vor Screeningvisite 1 erhielten, von der Studie ausgeschlossen wurden. Des Weiteren wurden Ausschlusskriterien ergänzt, um XLH-Patienten, die Denosumab in den letzten 6 Monaten vor Studienvisite 1 oder Teriparatid in den letzten 2 Monaten vor Studienvisite 1 einnahmen, von der Studie ausgeschlossen wurden. Diese Medikamente wurden auch auf die Liste der untersagten Medikamente während der Studie gesetzt.</li> <li>• Das Ausschlusskriterium #19 wurde modifiziert, sodass Erwachsene mit einer Vorgeschichte von immer wieder auftretenden XLH-bedingten Zahnabszessen an der Studie teilnehmen können.</li> </ul> <p><u>Änderung der Erhebungszeitpunkte:</u></p> <ul style="list-style-type: none"> <li>• Der Zeitpunkte der iFGF23-Messung, der Mutationsanalyse des <i>PHEX</i>-Gens sowie der Durchführung von Schwangerschaftstests wurde modifiziert bzw. spezifiziert. Der Mutationsstatus des <i>PHEX</i>-Gens wird nur einmal erhoben (entweder zu Studienvisite 2 oder zur Baseline-Visite).</li> <li>• Pharmakokinetische (PK) Messungen: Zusätzliche PK-Messungen werden in Woche 34, 36, 46 und 70</li> </ul>
--	--	--

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>durchgeführt. Radiographische Messungen erfolgen zur Baseline-Visite. Die Serumkonzentration von 1,25(OH)<sub>2</sub>D wird zusätzlich in Woche 24 analysiert.</p> <ul style="list-style-type: none"> <li>In der Fußnote 12 wurde angemerkt, dass der komplette WOMAC-Fragebogen erhoben wird.</li> </ul> <p><u>Behandlung:</u></p> <ul style="list-style-type: none"> <li>Es wurde spezifiziert, dass die subkutane Verabreichung der Prüfsubstanz entweder in den Bauch, die Oberarme oder die Oberschenkel erfolgen kann.</li> </ul> <p><u>Verblindung:</u></p> <ul style="list-style-type: none"> <li>Es wurde verdeutlicht, dass es sich um eine doppelblinde und Placebo-kontrollierte Studie handelt. Die Aufhebung der Verblindung aufgrund erhöhter Serumphosphat-Konzentrationen bezieht sich ausschließlich auf die doppelblinde Phase der Studie. Ein weiterer Abschnitt wurde hinzugefügt, der spezifiziert, dass die Verblindung des Prüfarztes bis Woche 48 aufrechterhalten wird.</li> </ul> <p><u>Endpunkte:</u></p> <ul style="list-style-type: none"> <li>Das Vorhandensein von Enthesopathien wurde als weiterer Wirksamkeitsendpunkt mit aufgenommen. Das Ausmaß der Enthesopathien wird anhand von Röntgenaufnahmen zu Baseline sowie zu Woche 24, 48 und 96 bestimmt.</li> <li>Es wurde spezifiziert, dass 2 Fragen des BPI-Fragebogens (Gebrauch von Schmerzmitteln und Schmerzlinderung durch Schmerzmittel) nicht erhoben werden, da diese separat mit einem Patiententagebuch analysiert werden.</li> </ul> <p><u>Generelle Durchführung:</u></p> <ul style="list-style-type: none"> <li>Es wurde hinzugefügt, dass bei einer unklaren Mutationsanalyse oder falls keine Mutation des <i>PHEX</i>-Gens vorliegt, weitere Mutationsanalysen von Genen erfolgen dürfen, die mit XLH assoziiert sind. Dies beinhaltet folgende Gene: <i>FGF23</i>, <i>Dentin Matrix Acidic Phosphoprotein 1 (DMP1)</i>, <i>Ectonucleotide Pyrophosphatase/Phosphodiesterase 1 (ENNP1)</i>, <i>Chloride Voltage-Gated</i></li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p><i>Channel 5 (CLCN5), und Solute Carrier Family 34 Member 3 (SLC34A3).</i></p> <p><u>Sicherheitsanalysen</u></p> <ul style="list-style-type: none"> <li>• Die Auswertung der Echokardiogramme erfolgt nicht im Studienzentrum, sondern zentral durch verblindete Personen.</li> <li>• Elektrokardiogramme werden als generelle Sicherheitsanalysen bei ektopischen Mineralisierungen durchgeführt.</li> <li>• Der Blutdruck wird zwei Mal bei Studienvisite gemessen.</li> <li>• Das Enzym Lipase wird bei allen Studienteilnehmern gemessen. Falls die Serumkonzentration der Amylase <math>\geq 1,5</math> x oberhalb der oberen Grenze des Normbereichs liegt, erfolgt eine reflexive Testung.</li> <li>• Die Beschreibung der FGF23-Messung wurde entfernt.</li> </ul> <p><u>Statistische Methoden:</u></p> <ul style="list-style-type: none"> <li>• Ein finale Analyse ist für Woche 96, dem Ende der Extensionsphase I, geplant.</li> <li>• Abschnitt 7.6.4.2 wurde dahingehend präzisiert, dass der wichtigste sekundäre Endpunkt sich aus den Mittelwerten des Patiententagebuches (tägliche Messung) sowie dem Messwert zur Studienvisite zusammensetzt.</li> <li>• In Abschnitt 7.6.4.3 wurde verdeutlicht, dass die Endpunkte <i>BPI Schmerzintensität</i> und <i>BPI Schmerzinterferenz</i> separat evaluiert werden. Der Endpunkt für BFI dagegen basiert auf der Frage 3 des BFI-Fragebogens. Die <i>BFI Schlimmste Fatigue</i> und die <i>BFI Globale Fatigue</i> berechnen sich aus dem Mittelwert der 9 Items des BFI-Fragebogens.</li> </ul> <p><u>Ethik:</u></p> <ul style="list-style-type: none"> <li>• Abschnitt 8.1.2 wurde aktualisiert und hinzugefügt, dass sowohl der Sponsor als auch der Prüfer alle Anstrengungen unternehmen, um sicherzustellen, dass die Studie vollständig nach den Prinzipien und den geltenden FDA-Vorschriften, den ICH-GCP-Richtlinien und lokalen</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>ethischen und regulatorischen Anforderungen durchgeführt wird.</p> <p><u>Aufbewahrung der Studienunterlagen:</u></p> <ul style="list-style-type: none"> <li>• Alle Studienunterlagen werden für mindestens 25 Jahre nach Studienende aufgehoben.</li> </ul> <p><u>Kontaktinformationen:</u></p> <ul style="list-style-type: none"> <li>• Die Informationen zum medizinischen Monitor wurden in Abschnitt 8.5.4.7 aktualisiert.</li> </ul> <p><u>Amendment 2 datiert auf den 08.09.2016:</u></p> <p><u>Erhebungszeitpunkte:</u></p> <ul style="list-style-type: none"> <li>• Die Fußnote der Übersichtstabelle zu den Erhebungszeitpunkten wurde aktualisiert. Ein doppelter Eintrag zu Studienvisite 34 wurde entfernt. Die Studienvisiten wurden neu nummeriert und eine nicht beabsichtigte Messung des Burosumab-AK zu Woche 70 entfernt.</li> </ul> <p><u>Ziele der Studie:</u></p> <ul style="list-style-type: none"> <li>• Die pharmakokinetischen Ziele der Studie wurden spezifiziert: die Pharmakokinetik von Burosumab wird während des Dossierungszyklus von der ersten Verabreichung bis zum Erreichen eines Steady-State-Levels analysiert.</li> </ul> <p><u>Genetische Untersuchungen:</u></p> <ul style="list-style-type: none"> <li>• Ein weiteres Gen (<i>Family with sequence similarity 20 (FAM20C)</i>) wurde in die Liste der Gene, bei denen Mutationsanalysen durchgeführt werden sollen, mitaufgenommen. Dies bezieht sich auf den Fall, dass keine <i>PHEX</i>-Mutation festgestellt werden kann oder die Mutationsanalyse nicht eindeutig ist.</li> </ul> <p><u>6MWT:</u></p> <ul style="list-style-type: none"> <li>• Abschnitt 7.5.4 wurde aktualisiert: der 6MWT wird zusätzlich zu Woche 72 bzw. zu einem früheren Zeitpunkt durchgeführt, falls ein Studienteilnehmer die Studie vor oder zu Woche 72 abbricht.</li> </ul> <p><u>Timed Up and Go-Test (TUG):</u></p> <ul style="list-style-type: none"> <li>• Abschnitt 7.5.5 wurde aktualisiert, um darauf hinzuweisen, dass der TUG-Test bei einem vorzeitigen Abbruch der Studie zwischen Woche 24 und Woche 72 zusätzlich bei der letzten Studienvisite durchgeführt wird.</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p><u>Konzentrationsmessung der Prüfsubstanz:</u></p> <ul style="list-style-type: none"> <li>Die Beschreibung der Konzentrationsmessung war ursprünglich in der Überschrift zum Abschnitt Enthesopathien enthalten und wurde nun im Abschnitt 7.5.7 eingefügt.</li> </ul> <p><u>Vitalparameter:</u></p> <ul style="list-style-type: none"> <li>Die Blutdruckmessung wurde spezifiziert. Diese erfolgt zu Studienvisite zwei Mal in einem Abstand von 15 min.</li> </ul> <p><u>Sekundäre Wirksamkeitsendpunkte:</u></p> <ul style="list-style-type: none"> <li>Abschnitt 7.6.4.3: Es wurde hinzugefügt, dass statistische Analysemethoden für zusätzliche sekundäre Wirksamkeitsendpunkte einschließlich der Adjustierung für multiples Testen im SAP beschrieben werden.</li> </ul> <p><u>Wirksamkeitsendpunkte:</u></p> <ul style="list-style-type: none"> <li>Die Beschreibung des Endpunkts <i>Heilen von Pseudofrakturen</i> wurde abgeändert in: <i>„Heilung von bereits existierenden Pseudofrakturen und/oder Looser-Umbauzonen, bestimmt anhand von Röntgenaufnahmen zu Baseline und anschließenden Röntgenaufnahmen.</i></li> </ul> <p><u>Amendment 3 datiert auf den 31.03.2017:</u></p> <p><u>Ziele der Studie:</u></p> <ul style="list-style-type: none"> <li>Neben dem primären Wirksamkeitsendpunkt, der sich auf Skelettschmerzen bezieht, wurden als weitere primäre Wirksamkeitsendpunkte der Effekt von Burosumab im Vergleich zu Placebo auf die Steifheit und physische Funktion definiert.</li> </ul> <p><u>Sekundäre Wirksamkeitsendpunkte:</u></p> <ul style="list-style-type: none"> <li>Es wurden zwei weitere sekundäre Wirksamkeitsendpunkte mit aufgenommen: Dies beinhaltet die Änderung von Baseline zu Woche 24 der Endpunkte <i>WOMAC Steifheit</i> und <i>WOMAC Physische Funktion</i>.</li> </ul> <p><u>Weitere sekundäre Wirksamkeitsendpunkte:</u></p> <ul style="list-style-type: none"> <li>Abschnitt 7.6.4.3 wurde aktualisiert: Es werden weitere sekundäre Wirksamkeitspunkte mit aufgenommen. Dies beinhaltet die</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Änderung des Endpunkts <i>BPI Schlimmster Schmerz</i> und <i>WOMAC Physische Funktion</i> von Baseline zu Post-Baseline-Visiten. Zusätzlich wurde klargestellt, dass sowohl die Änderung als auch die prozentuale Änderung der <i>Serumphosphat-Konzentration</i> zwischen Baseline und Post-Baseline-Visiten analysiert wird.</p> <p><u>Wirksamkeitsendpunkte:</u></p> <ul style="list-style-type: none"> <li>• Der Abschnitt 7.6.4.4 wurde aktualisiert: Der <i>Endpunkt Heilen von Frakturen und Looser-Umbauzonen</i> beinhaltet die Anzahl an aktiven Pseudofrakturen und/oder Frakturen zu Baseline sowie die Anzahl und den Prozentanteil an geheilten, teilweise geheilten, unveränderten oder sich verschlechternden (Pseudo-)Frakturen zu Post-Baseline-Visiten. Zusätzlich wird die Anzahl an Studienteilnehmern mit Veränderungen in der Anzahl an geheilten, teilweise geheilten, unveränderten oder sich verschlimmernden (Pseudo-)Frakturen, im Vergleich Baseline zu Post-Baseline, mit aufgenommen.</li> </ul> <p><u>Randomisierte Stratifizierung:</u></p> <ul style="list-style-type: none"> <li>• Die Randomisierung wird anhand der Schmerzintensität und der Region stratifiziert.</li> </ul> <p><u>Schwangerschaftstests und Verhütung:</u></p> <ul style="list-style-type: none"> <li>• Die Liste der hocheffektiven Verhütungsmethoden wurde aktualisiert.</li> </ul> <p><u>Studiendesign und Dauer:</u></p> <ul style="list-style-type: none"> <li>• Eine weitere Extensionsphase (Extensionsphase II) wurde in das Studienprotokoll mit aufgenommen, an der aber ausschließlich Studienteilnehmer aus den USA teilnehmen können. Diese ist über einen Zeitraum von bis zu 53 Wochen geplant und variiert je nach Studienteilnehmer.</li> <li>• Folgetelefonanrufe, um die Sicherheit zu analysieren, werden bis einschließlich 8 Wochen nach dem Beenden der Studie bzw. der finalen Studienvisite durchgeführt. Dies betrifft Studienteilnehmer, die Burosumab anschließend nicht kommerziell außerhalb der Studie</li> </ul>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>erhalten oder die Studie vorzeitig abbrechen. Die maximale Studiendauer wurde auf 157 Wochen verlängert. Für Studienteilnehmer außerhalb den USA endet die Studie jedoch nach 96 Wochen, gefolgt von Telefonanrufen für 8 Wochen, um die Sicherheit zu analysieren. Das Ende der Studie wird definiert als der Tag, an dem die letzte Untersuchung (einschließlich telefonischer Kontakt) für den letzten Studienteilnehmer der Studie durchgeführt wurde.</p> <p><u>Anwendung der Prüfsubstanz:</u></p> <ul style="list-style-type: none"> <li>• Abschnitt 7.4.1 und 7.4.6 wurden aktualisiert: Studienteilnehmer in den USA können nach ausreichendem Training durch das Studienpersonal die Prüfsubstanz eigenständig subkutan injizieren. Des Weiteren kann die Prüfsubstanz auch von geschultem Pflegepersonal zu Hause verabreicht werden. Zusätzliche Anweisungen bzgl. des Zeitpunkts des Trainings sowie die Umsetzung wurden ebenfalls hinzugefügt.</li> </ul> <p><u>Prüfer:</u></p> <ul style="list-style-type: none"> <li>• Dr. Karl Insogna wurde als leitender Prüfer auf die Titelseite mit aufgenommen.</li> </ul> <p><u>Amendment 4 datiert auf den 29.09.2017:</u> Die Änderungen sind nicht relevant, da diese nur die offene nicht-randomisierte Phase der Studie betreffen.</p> <p><u>Amendment 5 datiert auf den 26.01.2018:</u> Die Änderungen sind nicht relevant, da diese nur die offene nicht-randomisierte Phase der Studie betreffen.</p> <p><u>Amendment 1 des CSRW48 datiert auf den 05.06.2018</u> Die Änderungen sind nicht relevant, da diese nur die offene nicht-randomisierte Phase der Studie betreffen.</p>
4	Probanden / Patienten	<p>Anhand der statistischen Analysen war eine Studienpopulation von ca. 120 Probanden für die Rekrutierung geplant. Zunächst wurden 163 Erwachsene mit XLH gescreent und von diesen 134, anhand der Ein- und Ausschlusskriterien, in die Studie UX023-CL303 eingeschlossen. Alle 134 Studienteilnehmer erhielten während der doppelblinden Placebo-kontrollierten</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		Behandlungsphase von 24 Wochen mindestens eine Dosis von Burosumab oder Placebo und wurden sowohl in die Primär- als auch die Sicherheits- und Pharmakokinetischen Analysen mit einbezogen. Zusätzlich wurde der Zeitraum der Screeningphase auf > 31 Tage abgeändert.
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p><b>Einschlusskriterien:</b></p> <p>Erwachsene, die in die Studie mit aufgenommen wurden, mussten folgende Einschlusskriterien erfüllen:</p> <ul style="list-style-type: none"> <li>• Männer und Frauen im Alter von 18 – 65 Jahren</li> <li>• Krankheitsbefund von XLH anhand der klassischen Kriterien wie z. B. Kleinwuchs und Knochendeformationen <u>und mindestens</u> einem weiteren klinischen oder biochemischen XLH-spezifischen Merkmal <ul style="list-style-type: none"> <li>○ <i>PHEX</i>-Mutation oder direktes Familienmitglied mit entsprechender X-chromosomaler Vererbung</li> <li>○ Serumkonzentration des intakten Fibroblasten-Wachstumsfaktors 23 (FGF23) &gt; 30 pg/ml (analysiert mit dem Kainos-Assay)</li> </ul> </li> <li>• Biochemische Befunde bei der zweiten Screeningvisite, die ebenfalls mit XLH assoziiert sind <ul style="list-style-type: none"> <li>○ <i>Serumphosphat-Konzentration</i> &lt; 2,5 mg/dl</li> <li>○ <i>TmP/GFR</i> &lt; 2,5 mg/dl (Quotient der maximalen tubulären Rückresorption von Phosphat (TmP) und der glomerulären Filtrationsrate (GFR))</li> </ul> </li> <li>• Skelettschmerzen, die auf die XLH-Erkrankung und Osteomalazie zurückzuführen sind (<i>BPI Schlimmster Schmerz</i> ≥ 4 bei der ersten Studienvisite)</li> <li>• Geschätzte glomeruläre Filtrationsrate eGFR ≥ 60 ml/min (<i>Chronic kidney disease epidemiology collaboration</i> (CDK-EPI)) oder eGFR von 45 &lt; 60 ml/min bei der zweiten Studienvisite und Bestätigung, dass</li> </ul>

		<p>die Niereninsuffizienz nicht durch eine Nephrokalzinose bedingt ist</p> <ul style="list-style-type: none"> <li>• Bei Einnahme von Schmerzmitteln (narkotische Analgetika, Opioide) zur Behandlung von chronischen Schmerzen, muss für mindestens 21 Tage vor Screeningvisite 1 eine gleichbleibende Dosierung eingehalten werden. Diese Dosis darf auch während der Placebo-kontrollierten Behandlungsphase der Studie nicht geändert werden. Zudem soll eine Dosis von 60 mg/Tag an Morphin (oder einem Äquivalent) nicht überschritten werden.</li> <li>• Schriftliche Einwilligung nach Erläuterung der Studie</li> <li>• Zugang zu früheren medizinischen Aufzeichnungen für die Erfassung von biochemischen- und röntgenologischen Daten und dem Krankheitsverlauf</li> <li>• Frauen im gebärfähigen Alter müssen während des Screenings einen negativen Schwangerschaftstest aufweisen und bereit sein, auch während der Studie regelmäßig Schwangerschaftstests durchführen zu lassen</li> <li>• Frauen, die als nicht gebärfähig klassifiziert sind, müssen sich bereits mindestens 2 Jahre vor Studienbeginn in der Menopause befunden haben oder sich mindestens 1 Jahr vor dem Screening einer Eileiterunterbindung oder einer bilateralen Salpingo-Oophorektomie unterzogen haben</li> <li>• Studienteilnehmer im gebärfähigen Alter oder diejenigen, die Partner im gebärfähigen Alter haben, die sich nicht einer vollständigen Hysterektomie oder bilateralen Salpingo-Oophorektomie unterzogen haben und sexuell aktiv sind, müssen bereit sein, zwei effektive Verhütungsmethoden (orale hormonelle Kontrazeptiva, hormonelle Verhütungspflaster, Vaginalring, Intrauterinpeessar, physikalische Doppel-Barrieremethoden, chirurgische Hysterektomie, Vasektomie, Tubenligatur oder Abstinenz) nach Unterschreibung der Einwilligungserklärung über die gesamte Dauer der Studie sowie</li> </ul>
--	--	---

		<p>einschließlich 12 Wochen nach Absetzen der Studienmedikation anzuwenden</p> <ul style="list-style-type: none"> <li>• Der Patient muss nach Ansicht des Prüfarztes bereit und in der Lage sein, alle Aspekte der Studie und den Studienvisiten-Plan einzuhalten und den Anweisungen Folge zu leisten</li> <li>• Studienteilnehmer, die in die Studie mit aufgenommen werden, müssen mindestens 4 von 7 Tagen des Patiententagebuchs vor Baseline-Visite abgeschlossen haben</li> </ul> <p><b>Ausschlusskriterien:</b></p> <p>Erwachsene, die mindestens eines der folgenden Kriterien erfüllten, durften nicht an der Studie teilnehmen:</p> <ul style="list-style-type: none"> <li>• Verwendung von Vitamin D-Metaboliten (Calcitriol, Doxercalciferol oder Paricalcitol) innerhalb der letzten 14 Tage vor Screeningvisite 2</li> <li>• Verwendung von oralem Phosphat innerhalb der letzten 14 Tage vor Screeningvisite 2</li> <li>• Verwendung von Aluminiumhydroxid-Antazida, Acetazolamid und Thiaziden innerhalb der letzten 7 Tage vor Screeningvisite 2</li> <li>• Langfristige Einnahme von systemischen Kortikosteroiden (&gt; 10 Tage innerhalb der letzten 2 Monate vor Screeningvisite 1)</li> <li>• Serumkalziumkonzentration <math>\geq 10,8</math> mg/dl zu Screeningvisite 2</li> <li>• Serumkonzentration des intakten Parathormons (iPTH) <math>\geq 2,5</math> <i>upper limit of normal</i> (ULN) zu Screeningvisite 1</li> <li>• Verwendung von Medikamenten zur Reprimierung des Parathormons (z. B. Cinacalcet) innerhalb 60 Tage vor Screeningvisite 1</li> <li>• Verwendung von Biphosphonaten innerhalb der letzten 2 Jahre vor Screeningvisite 1</li> <li>• Verwendung von Denosumab innerhalb der letzten 6 Monate vor Screeningvisite 1</li> <li>• Verwendung von Teriparatid innerhalb der letzten 2 Monate vor Screeningvisite 1</li> </ul>
--	--	---

		<ul style="list-style-type: none"> <li>• Geplanter oder empfohlener orthopädischer Eingriff innerhalb der ersten 24 Wochen der klinischen Studie</li> <li>• Traumatische Frakturen oder orthopädische Eingriffe innerhalb der letzten 6 Monate vor Screeningvisite 1</li> <li>• Verwendung von KRN23 oder anderen therapeutischen monoklonalen Antikörpern innerhalb der letzten 90 Tage vor Screeningvisite 1</li> <li>• Verwendung des Prüfprodukts oder des Prüfmedikaments innerhalb der letzten 30 Tage vor Screeningvisite 1 oder Bedarf eines Prüfprodukts vor Abschluss aller geplanten Studienbewertungen oder in <u>Japan</u> die Verwendung eines Prüfprodukts oder eines Prüfmedikaments innerhalb der letzten 4 Monate vor Screeningvisite oder Bedarf der Anwendung eines Prüfpräparats vor Abschluss aller Studienauswertungen</li> <li>• Schwangerschaft oder Stillen zu Beginn der Screening/Baseline-Visite oder die Beabsichtigung während der Laufzeit der Studie schwanger zu werden (bezieht sich entweder auf den Studienteilnehmer oder Partner)</li> <li>• Nicht fähig oder nicht willens, auf die in der Studie nicht erlaubten Medikamente während der gesamten Laufzeit der Studie zu verzichten</li> <li>• Vorhandensein oder Anamnese einer Hypersensitivität, Allergien oder anaphylaktischen Reaktionen auf monoklonale Antikörper oder Burosumab, wodurch der Studienteilnehmer laut Prüfarzt einem erhöhten Risiko von unerwünschten Ereignissen ausgesetzt ist</li> <li>• Vorgeschichte von folgenden positiven Tests: HIV, Hepatitis-B Oberflächenantigen und/oder Hepatitis-C-Antikörper</li> <li>• Vorgeschichte von rezidivierenden Infekten (außer Zahnabszesse die mit XLH assoziiert sind), sowie Infektionsanfälligkeit oder Immunschwäche</li> </ul>
--	--	--

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> <li>• Vorgeschichte von bösartigen Neoplasien (ausgenommen Basalzellkarzinome)</li> <li>• Vorliegen einer Erkrankung oder eines Zustandes, der mit der Studienteilnahme oder Sicherheit interferieren würde</li> <li>• Vorhandensein oder Anamnese einer Erkrankung, die laut Prüfarzt, das Risiko der Nichteinhaltung der Behandlung oder des vorzeitigen Abbrechens der Studie erhöht</li> </ul>
4b	Studienorganisation und Ort der Studiendurchführung	<p>Sponsor war Ultragenyx Pharmaceutical Inc. Novato, CA, USA</p> <p>Insgesamt wurden 134 Erwachsene im Alter zwischen 18 – 65 Jahren in 25 Studienzentren in 7 Ländern aufgenommen (8 in den USA, 5 im Vereinigten Königreich Großbritannien, 5 in Japan, 3 in Frankreich, 2 in Südkorea, 1 in Irland, und 1 in Italien)</p>
5	<p>Interventionen</p> <p>Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.</p>	<p>Burosumab ist ein humaner monoklonaler Antikörper, der den im Überschuss vorhandenen Fibroblasten-Wachstumsfaktor 23 (FGF23) bindet und dadurch dessen Aktivität hemmt und die Serumphosphat-Homöostase langfristig wiederherstellt.</p> <p>Je 1 ml Burosumab wurde als sterile klare Lösung ohne Konservierungsstoffe in 5 ml-Einweg-Durchstechflaschen mit einer Konzentration von 30 mg/ml zur Verfügung gestellt.</p> <p>Die Dosierung von Burosumab war abhängig vom Körpergewicht (KG) der Studienteilnehmer. Insgesamt wurde je kg KG eine Dosis von 1 mg Burosumab subkutan verabreicht. Die Maximaldosis betrug 90 mg. Die Dosis wurde über die gesamte Studiendauer beibehalten, außer wenn die <i>Serumphosphat-Konzentration</i> zu irgendeinem Zeitpunkt 5 mg/dl überstieg oder zweimal einen Wert von über 4,5 mg/dl erreichte. Eine Dosisanpassung fand zudem statt, falls sich das Gewicht der Studienteilnehmer um mehr als 20 % zum Baseline-Wert veränderte. In diesem Fall wurde die notwendige Burosumab-Konzentration neu berechnet und dem veränderten Körpergewicht angepasst.</p> <p>Für den Fall, dass die <i>Serumphosphat-Konzentration</i> 5 mg/dl überstieg, wurde die Verblindung des Prüfarztes</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>aufgehoben und die Burosumab-Dosis auf die Hälfte reduziert.</p> <p>Falls die <i>Serumphosphat-Konzentration</i> über 4,5 mg/dl anstieg, aber unterhalb von 5 mg/dl blieb, wurde die Verblindung des Prüfarztes ebenfalls aufgehoben. Die Burosumab-Dosis wurde jedoch nur dann auf die Hälfte reduziert, falls ein zweites Mal eine Konzentration von über 4,5 mg/dl überschritten wurde.</p> <p>Burosumab wurde subkutan in Bauch, Oberarme oder Beine injiziert, wobei die Injektionsstelle bei jeder Verabreichung gewechselt wurde. Die maximale Menge, die je Injektionsstelle verabreicht werden konnte, betrug 1,5 ml. Falls ein größeres Volumen notwendig war, wurde Burosumab an mehreren Stellen injiziert, wobei immer eine andere Injektionsstelle verwendet wurde.</p> <p><u>Burosumab-Gruppe:</u></p> <p>Die Dosis von Burosumab betrug 1 mg/kg KG und wurde alle 4 Wochen (28 Tage) s.c. verabreicht.</p> <p><u>Placebo-Gruppe:</u></p> <p>Je 1 ml Placebo wurde als sterile klare Lösung ohne Konservierungsstoffe zur Verfügung gestellt. Die Zusammensetzung war identisch zu der von Burosumab, mit dem einzigen Unterschied, dass die Placebo-Lösung nicht die aktive Substanz enthielt. Die Injektion erfolgte wie zuvor für Burosumab beschrieben.</p>
<b>6</b>	Zielkriterien	
<b>6a</b>	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p><u>Wirksamkeit:</u></p> <p>Für die Nutzenbewertung ist nur die doppelblinde Phase der Studie (bis Woche 24) relevant. In dieser doppelblinden Phase waren sowohl die Studienteilnehmer, als auch das Studienpersonal sowie der Prüfarzt gegenüber der Gruppenzuteilung und Behandlung verblindet.</p> <p><b>Primärer Wirksamkeitsendpunkt:</b></p> <p>Anteil an Studienteilnehmern, die in der Mitte des Dosisintervalls eine <i>Serumphosphat-Konzentration</i> oberhalb der unteren Grenze des Normbereichs von 2,5 mg/dl erreichten, gemittelt über die</p>

		<p>Dosiszyklen zwischen Baseline und Woche 24</p> <ul style="list-style-type: none"> <li>• Erhebungszeitpunkte: Woche 2, 6, 10, 14, 18, 22, 24</li> <li>• Aufgrund der fluktuierenden Serumphosphat-Konzentrationen, die mit dem circadianen Rhythmus assoziiert sind, wurden die Blutproben morgens möglichst jeweils zur gleichen Uhrzeit entnommen.</li> <li>• Die Auswertung erfolgte verblindet durch ein zentrales Labor.</li> </ul> <p><b>Sekundäre Wirksamkeitsendpunkte:</b> Evaluation des Effekts von Burosumab im Vergleich zu Placebo auf folgende wesentliche sekundäre Endpunkte:</p> <ul style="list-style-type: none"> <li>• Änderung des Endpunkts <i>BPI Schlimmster Schmerz</i> von Baseline zu Woche 12, 24, 36, 48, 72, 96 <ul style="list-style-type: none"> <li>○ Validierung durch Interviewstudie und Studie UX023-CL203</li> </ul> </li> <li>• Änderung des Endpunkts <i>WOMAC Steifheit</i> von Baseline zu Woche 12, 24, 36, 48, 72, 96 <ul style="list-style-type: none"> <li>○ Validierung durch Interviewstudie und Studie UX023-CL203</li> </ul> </li> <li>• Änderung des Endpunkts <i>WOMAC Physische Funktion</i> von Baseline zu Woche 12, 24, 36, 48, 72, 96 <ul style="list-style-type: none"> <li>○ Validierung durch Interviewstudie und Studie UX023-CL203</li> </ul> </li> </ul> <p><b>Weitere sekundäre Wirksamkeitsendpunkte:</b></p> <ul style="list-style-type: none"> <li>• Zusätzliche Messmethoden zur Evaluation der <i>Serumphosphat-Konzentration</i> (zu Baseline und Woche 1, 2, 6, 10, 14, 18, 22, 24, 26, 28, 34, 36, 46, 48, 60, 70, 72, 84, 94, 96 und alle 12 Wochen in der Extensionsphase II <math>\leq</math> 149): <ul style="list-style-type: none"> <li>○ Anteil der Patienten, die am Ende des Dosiszyklus (4 Wochen nach der Dosierung) eine mittlere <i>Serumphosphat-Konzentration</i> über der unteren Grenze des Normbereichs erreichten, gemittelt über die Dosiszyklen</li> </ul> </li> </ul>
--	--	--

		<ul style="list-style-type: none"> <li>○ Mittlere Änderung von Baseline und prozentuale Änderung von Baseline bis zur Mitte jedes Dosiszyklus bzw. am Ende des Dosiszyklus, gemittelt über die Dosiszyklen</li> <li>○ Zeitabhängige <i>Area Under the Curve</i> (AUC)</li> <li>○ Aufgrund der fluktuierenden Serumphosphat-Konzentrationen, die mit dem circadianen Rhythmus assoziiert sind, wurden die Blutproben morgens möglichst jeweils zur gleichen Uhrzeit entnommen.</li> <li>○ Die Auswertung erfolgte verblindet durch ein zentrales Labor.</li> <li>● Pharmakodynamische Marker, die den Zustand der Phosphat-Homöostase über den Studienverlauf reflektieren       <ul style="list-style-type: none"> <li>○ <i>Phosphat</i> im Serum und Urin (zu Baseline und Woche 1, 2, 6, 10, 14, 18, 22, 24, 26, 28, 34, 36, 46, 48, 60, 70, 72, 84, 94, 96 und alle 12 Wochen in der Extensionsphase II <math>\leq 149</math>)</li> <li>○ Serum-<i>1,25(OH)<sub>2</sub>-Vitamin D<sub>3</sub></i> (Baseline, Woche 1, 24, 20, 21, 22, 24, 46, 48, 60, 70, 72, 84, 94, 96 und alle 12 Wochen in der Extensionsphase II <math>\leq 149</math>)</li> <li>○ <i>TmP/GFR</i> (Baseline, Woche 2, 4, 12, 22, 24, 48, 60, 72, 84, 96 und <math>\leq 149</math>)</li> <li>○ <i>TRP</i> (Baseline, Woche 2, 4, 12, 22, 24, 48, 60, 72, 84, 96 und <math>\leq 149</math>)</li> </ul> </li> <li>● Änderung und prozentuale Änderung von Baseline zu Post-Baseline in biochemischen Parametern des Knochenstoffwechsels einschließlich <i>Procollagen Type I N-Propeptid (PINP)</i>, <i>Carboxy-terminales Typ-I-Kollagen Telopeptid (CTx)</i> und die <i>knochenspezifische alkalische Phosphatase (BALP)</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72, 96 und <math>\leq 149</math></li> </ul> </li> </ul>
--	--	--

		<ul style="list-style-type: none"> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>BPI Schlimmster Schmerz</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72 und 96</li> </ul> </li> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>BPI Schmerzintensität</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72 und 96</li> </ul> </li> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>BPI Schmerzinterferenz</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72 und 96</li> </ul> </li> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>Brief Fatigue Inventory (BFI) Schlimmste Fatigue</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72 und 96</li> </ul> </li> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>BFI Globale Fatigue</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72 und 96</li> <li>○</li> </ul> </li> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>WOMAC Steifheit</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72, 96 und alle 24 Wochen in der Extensionsphase II <math>\leq 149</math></li> </ul> </li> <li>• Änderung von Baseline zu Post-Baseline des Endpunkts <i>WOMAC Physische Funktion</i> <ul style="list-style-type: none"> <li>○ Erhebungszeitpunkte: Baseline, Woche 12, 24, 36, 48, 72, 96 und alle 24 Wochen in der Extensionsphase II <math>\leq 149</math></li> </ul> </li> </ul> <p><b>Explorative Wirksamkeitsendpunkte:</b></p> <ul style="list-style-type: none"> <li>• Änderung der <i>Anzahl und des Prozentsatzes an aktiven Pseudo-frakturen/Frakturen</i>, ermittelt anhand von Röntgenbildaufnahmen, die im Vergleich zu Baseline zu Post-</li> </ul>
--	--	---

		<p>Baseline geheilt waren, unverändert oder sich verschlechtert hatten</p> <ul style="list-style-type: none"> <li>○ Erhoben zu Baseline, Woche 12, 24, 36 und 48 sowie bei Klinikvisiten, wenn eine neue Fraktur diagnostiziert wurde</li> <li>○ Schulung und Verblindung der zwei zentralen Auswerter</li> </ul> <ul style="list-style-type: none"> <li>• Änderung der <i>Anzahl an Studienteilnehmern mit aktiven Pseudo-frakturen/Frakturen</i>, ermittelt anhand von Röntgenbildaufnahmen, die im Vergleich zu Baseline zu Post-Baseline geheilte oder unveränderte Frakturen aufwiesen oder deren Frakturen sich verschlechtert hatten <ul style="list-style-type: none"> <li>○ Erhoben zu Baseline, Woche 12, 24, 36 und 48 sowie bei Klinikvisiten, wenn eine neue Fraktur diagnostiziert wurde</li> <li>○ Schulung und Verblindung der zwei zentralen Auswerter</li> </ul> </li> <li>• Änderung der <i>Gesamtbelastung der Fersenbeine</i> von Baseline zu Post-Baseline aufgrund von Enthesopathien (Summe der Dimensionen der oberen und unteren Fersenspornbeine, bilateral in zwei Dimensionen gemessen), ermittelt anhand von seitlichen Röntgenbildaufnahmen des Fußes <ul style="list-style-type: none"> <li>○ Erhoben zu Baseline, Woche 24, 48, 96</li> <li>○ Schulung und Verblindung der zwei zentralen Auswerter</li> </ul> </li> <li>• Änderungen des Endpunkts <i>Patient Global Impression of Improvement (PGI-I)</i> von Baseline zu Post-Baseline <ul style="list-style-type: none"> <li>○ Erhoben zu Baseline, Woche 12, 24, 36 und 48</li> </ul> </li> <li>• Veränderung der Gehfähigkeit anhand des <i>6MWT</i> im Vergleich Baseline zu Post-Baseline <ul style="list-style-type: none"> <li>○ Erhoben zu Baseline, Woche 12, 24, 36, 48, 72, 96 und alle 24 Wochen in der Extensionsphase II <math>\leq 149</math></li> <li>○ Schulung des klinischen Personals auf Basis der ATS Guidelines</li> </ul> </li> <li>• Veränderung der Beweglichkeit anhand des <i>Timed Up and Go-Test</i></li> </ul>
--	--	---

		<p>(TUG) im Vergleich zu darauffolgenden Studienvisiten</p> <ul style="list-style-type: none"> <li>○ Erhoben zu Woche 24, 36, 48, 72, 96 und alle 24 Wochen in der Extensionsphase II <math>\leq</math> 149</li> </ul> <p><b>Pharmakokinetische Endpunkte:</b></p> <ul style="list-style-type: none"> <li>• Serumkonzentration von <i>Burosumab</i> während der Exposition und beim Erreichen des Spitzenwertes <ul style="list-style-type: none"> <li>○ Erhoben zu Woche 1, 2, 4, 21, 22, 24, 34, 36, 46, 48, 70, 72 und 96</li> </ul> </li> </ul> <p><b>Sicherheitsendpunkte:</b></p> <ul style="list-style-type: none"> <li>• Auftreten, Häufigkeit und Schwere der <i>unerwünschten Ereignisse (UEs)</i> und <i>schwerwiegende unerwünschten Ereignisse (SUEs)</i> <ul style="list-style-type: none"> <li>○ Baseline, Woche 1-96 und alle 4 Wochen in der Extensionsphase II <math>\leq</math> 149</li> </ul> </li> <li>• <i>Frühere und begleitende Medikation</i> <ul style="list-style-type: none"> <li>○ Baseline, Woche 1-96 und alle 4 Wochen in der Extensionsphase II <math>\leq</math> 149</li> </ul> </li> <li>• Änderung von <i>chemischen, hämatologischen und Urinkontrollparametern</i> von Baseline zu Post-Baseline <ul style="list-style-type: none"> <li>○ Baseline, Woche 4, 12, 24, 36, 48, 60, 72, 84, 96 und alle 12 Wochen in der Extensionsphase II <math>\leq</math> 149</li> </ul> </li> <li>• Änderung der Vitalzeichen und des <i>Gewichts</i> im Vergleich Baseline zu Post-Baseline <ul style="list-style-type: none"> <li>○ Baseline, Woche 12, 24, 36, 48, 72, 96, 108 und alle 12 Wochen in der Extensionsphase II (<math>\leq</math> 149)</li> </ul> </li> <li>• Änderung der <i>geschätzten glomerulären Filtrationsrate (eGFR)</i> im Vergleich Baseline zu Post-Baseline <ul style="list-style-type: none"> <li>○ Baseline, Woche 2, 4, 12, 22, 24, 48, 60, 72, 84, 96 und <math>\leq</math> 149</li> </ul> </li> <li>• Änderung des <i>Echokardiogramm (ECHO)</i> von Baseline zu Post-Baseline <ul style="list-style-type: none"> <li>○ Baseline, Woche 24, 48, 72, 96 und <math>\leq</math> 149</li> </ul> </li> </ul>
--	--	--

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> <li>• Änderung des <i>Elektrokardiogramms (EKG)</i> von Baseline zu Post-Baseline               <ul style="list-style-type: none"> <li>○ Baseline, Woche 24, 48, 72, 96 und <math>\leq 149</math></li> </ul> </li> <li>• Veränderung des <i>Nephrokalzinose-Scores</i> von Baseline zu Post-Baseline ermittelt anhand eines Nierultraschalls               <ul style="list-style-type: none"> <li>○ Woche 24, 48, 72, 96 und <math>\leq 149</math></li> </ul> </li> <li>• <i>Körperliche Untersuchungen</i> <ul style="list-style-type: none"> <li>○ Baseline, Woche 12, 24, 36, 48, 72, 96, 108 und <math>\leq 149</math></li> </ul> </li> <li>• Änderung des <i>Anti-Burosumab-AK-Status</i> von Baseline zu Post-Baseline               <ul style="list-style-type: none"> <li>○ Woche 4, 24, 48, 72, 96 und <math>\leq 149</math></li> </ul> </li> </ul>
<b>6b</b>	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Die explorativen Wirksamkeitsendpunkte <i>WOMAC Steifheit</i> und <i>WOMAC Physische Funktion</i> wurden nachträglich als sekundäre Wirksamkeitsendpunkte klassifiziert. Dies erfolgte, da sowohl muskuloskeletale Schmerzen als auch Steifheit die Mobilität und die physische Funktion von XLH-Patienten erheblich beeinflussen. Daher wurde entschieden, dass die klinische Relevanz von Burosumab am besten durch Erhebung der drei Endpunkte <i>BPI Schlimmster Schmerz</i> , <i>WOMAC Steifheit</i> und <i>WOMAC Physische Funktion</i> evaluiert werden kann.
7	Fallzahl	Etwa 120 Probanden waren für die Aufnahme in die Studie geplant. 134 Patienten wurden eingeschlossen, die alle mindestens eine Dosis der Studienmedikation während der Placebo-kontrollierten Behandlungsphase erhalten haben.
<b>7a</b>	Wie wurden die Fallzahlen bestimmt?	<p>Ziel der Phase III-Studie war es, die Wirksamkeit von Burosumab hinsichtlich des Erreichens einer <i>mittleren Serumphosphat-Konzentration</i> oberhalb der unteren Grenze des Normbereichs von 2,5 mg/dl in der Mitte des Dosisintervalls zu evaluieren.</p> <p>Es wurde angenommen, dass nach 24 Wochen 60 % der Studienteilnehmer, die Burosumab erhielten und 10 % der Studienteilnehmer die stattdessen Placebo verabreicht bekommen hatten, eine <i>mittlere Serumphosphat-Konzentration</i> von <math>\geq 2,5</math> mg/dl erreichten. Unter dieser</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Annahme wurde ermittelt, dass eine Gesamtstichprobengröße von 120 (60 Studienteilnehmer je Behandlungsgruppe) ausreichend ist, um zum zweiseitigen Signifikanzniveau von 0,05 eine statistische Güte (Power) von &gt; 95 % für die Identifikation des angenommenen Unterschiedes von 50 Prozentpunkten zu erreichen.</p> <p>Aufgrund der Anzahl von 120 Studienteilnehmern konnte darüber hinaus mit einer Power von <math>\geq 80\%</math> eine Veränderung von 1,0 zu Baseline bei dem Endpunkt <i>BPI Schlimmster Schmerz</i> detektiert werden. Hierbei wurde eine mittlere Veränderung von 2,0 zu Baseline in der Burosumab-Gruppe sowie eine Veränderung von 1,0 in der Placebo-Gruppe angenommen mit einer Standardabweichung von 1,8 und einer Drop-Out-Rate von 10 %.</p> <p>Insgesamt wurden 134 Patienten in die Studie mit aufgenommen und 1:1 auf die zwei Behandlungsgruppen (Burosumab (n = 68) und Placebo (n = 66) randomisiert.</p>
<b>7b</b>	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	<p>Die Primäranalyse erfolgte zu Woche 24 nachdem alle Studienteilnehmer die Placebo-kontrollierte Phase beendet hatten.</p> <p>Alle Studienteilnehmer, außer einer aus der Burosumab-Gruppe, beendeten die 24-wöchige Studienphase inklusive der letzten Studienvisite. Der einzige Studienteilnehmer, der die Studie vorzeitig verlassen hatte, erhielt zwar die letzte Burosumab-Injektion 4 Wochen vor Studienabbruch, nahm jedoch nicht an der letzten Studienvisite zu Woche 24 teil.</p> <p>Um die langfristige Wirkung und Sicherheit von Burosumab zu untersuchen, erfolgte eine weitere Analyse nach 48 Wochen (Behandlungsfortsetzungsphase, offenes Studiendesign, Wechsel der Placebo-Gruppe in die Burosumab-Gruppe) nachdem alle Studienteilnehmer die Bewertung zu Woche 48 abgeschlossen oder während der Extensionsphase die Studie abgebrochen hatten.</p> <p>Anschließend erfolgte eine weitere 48-wöchige Extensionsphase I bei der alle Studienteilnehmer, wie bereits zuvor beschrieben, Burosumab erhielten. Für die</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Studienteilnehmer aus Europa, Japan und Südkorea war die Studie nach 96 Wochen beendet. Studienteilnehmer aus den USA hatten dagegen noch die Möglichkeit an einer weiteren Extensionsphase (Extensionsphase II, maximal 53 Wochen) teilzunehmen.</p> <p>Administrative Analysen konnten während der Behandlungsfortsetzungsphase (Woche 24 – 48), der Extensionsphase I (Woche 48 – 96) und Extensionsphase II (Woche 96 – ≤ 149, nur in den USA) durchgeführt werden, um regulatorische Fragestellungen zu klären. Eine Zwischenanalyse zur Untersuchung der Wirksamkeit war bei dieser Studie nicht geplant.</p>
<b>8</b>	Randomisierung, Erzeugung der Behandlungsfolge	Siehe Antwortfeld 8 a und 8 b
<b>8a</b>	Methode zur Generierung der zufälligen Zuteilung	<p>Patienten, die die Einschlusskriterien erfüllten und in die Studie aufgenommen wurden, erhielten vor Studienbeginn eine Identifikationsnummer. Anschließend wurden die Studienteilnehmer anhand eines <i>Interaktiven Web Response-System</i> (IWRS) 1:1 entweder der Burosumab- oder der Placebogruppe zugeteilt [52]. Die Randomisierung wurde mit dem Patiententagebuchbericht „<i>BPI Durchschnittlicher Schmerz</i>“ stratifiziert. Zusätzlich erfolgte eine Stratifizierung nach Region (Nordamerika/EU, Japan und Südkorea), um das Gleichgewicht zwischen den beiden Behandlungsgruppen in Japan bzw. in Südkorea zu gewährleisten, da nur wenige Teilnehmer erwartet wurden.</p> <p>Der Randomisierungsplan wurde von einem unabhängigen Drittanbieter erstellt [52].</p> <p>Ebenfalls durften nicht mehr als 70 % an weiblichen Studienteilnehmer aufgenommen werden.</p>
<b>8b</b>	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	<p>Die Randomisierung wurde anhand der Schmerzintensität, <i>BPI Durchschnittlicher Schmerz</i>, basierend auf einem Patiententagebuch, das in einem Zeitraum über 7 Tage vor der Visite zu Baseline geführt wurde, stratifiziert. Nur Studienteilnehmer, die das Patiententagebuch für mindestens 4 Tage geführt hatten, wurden auch tatsächlich in die Studie aufgenommen. Ursprünglich sollte die Stratifizierung anhand des <i>BPI</i></p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p><i>Schlimmsten Schmerzes</i> innerhalb der letzten 24 Stunden (Frage 3 des <i>BPI</i> Fragebogens, &gt; 6,0 oder ≤ 6,0) erfolgen. Aufgrund eines Fehlers des IWRS erfolgte die Stratifizierung jedoch nach Frage 5 <i>BPI Durchschnittlicher Schmerz</i> (&gt; 6,0 oder ≤ 6,0). Ursache hierfür war, dass das Patiententagebuch nicht die Frage 1 und 2 des ursprünglichen Schmerzfragebogens enthielt und daher die Frage 3 <i>BPI Schlimmster Schmerz</i> als Frage 1 deklariert wurde und die Frage 5 <i>BPI Durchschnittlicher Schmerz</i> als Frage 3. Das IWRS stratifizierte anschließend die Probanden nach Frage 3 <i>BPI Durchschnittlicher Schmerz</i>. Da die Rekrutierung der Patienten und Verabreichung der Studiensubstanz bereits begonnen hatte als der Fehler entdeckt wurde, wurde beschlossen, dass die Stratifizierung weiterhin anhand der Frage 3 <i>BPI Durchschnittlicher Schmerz</i> erfolgen sollte, anstatt die Stratifizierung in der Mitte der Studie zu ändern. Der IWRS-Fehler wurde als Protokollabweichung im elektronischen Datenerfassungssystem (EDC) aufgezeichnet.</p> <p>Ein weiterer Stratifizierungsfaktor, der mit dem Amendment 3 in das Studienprotokoll aufgenommen wurde, war der Faktor Region. Hierbei ist jedoch hervorzuheben, dass bereits beim Einschluss des ersten Patienten die Randomisierung nach Schmerz und Region stratifiziert war und in der Randomisierungsspezifikation hinterlegt ist [52]. Der Stratifizierungsfaktor Region wurde aufgrund von operativen und logistischen Gründen zusätzlich zu dem Merkmal Schmerz aufgenommen. Da lediglich eine geringe Anzahl an Studienteilnehmern aus Japan und Korea zu erwarten war, sollte dies eine Gleichverteilung auf die beiden Behandlungsgruppen Burosumab und Placebo gewährleisten. (Nordamerika/EU, Japan und Südkorea).</p>
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Bei der Studie UX023-CL303 handelt es sich um eine doppelblinde Studie. Weder dem Sponsor der Studie noch den Studienteilnehmern oder Personen, die an der Ausführung der Studie beteiligt waren, war die Zuteilung der Studienteilnehmer in

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>die verschiedenen Behandlungsgruppen bekannt.</p> <p>Zur Verblindung wurde den Studienteilnehmern eine Nummer zugeteilt. Weder dem Studienpersonal noch den Personen vor Ort war die Zuteilung der Nummern (Randomisierungscode) sowie die Behandlung der Studienteilnehmer bekannt. Die Verblindung konnte nur im Falle von Sicherheitsaspekten aufgehoben werden (Entblindung Prüfarzt). Die Prüfsubstanz wurde mit der Studiennummer und einer spezifischen Kit-Nummer versehen. Die Lieferung der Prüfsubstanzen zu den Studienzentren erfolgte ebenfalls so, dass die Verblindung aufrechterhalten werden konnte.</p> <p>Beide Prüfsubstanzen, Placebo und Burosumab, hatten darüber hinaus das gleiche Erscheinungsbild.</p> <p>Für alle Messungen von Serum und Urin wurden zentrale Labors verwendet und die Messwerte weder dem Sponsor noch dem Studienpersonal während der doppelblinden Studienphase mitgeteilt.</p> <p>Röntgenbildaufnahmen, Echokardiogramme, Nierenultraschalluntersuchungen und Elektrokardiographien wurden zentral ausgewertet, wobei das Personal verblindet gegenüber der Behandlung und den Studienteilnehmerdaten war.</p>
<b>10</b>	<p>Randomisierung, Durchführung</p> <p>Wer hat die Randomisierungsliste erstellt, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?</p>	<p>Die Randomisierungsliste wurde mittels einem IWRS, welches von einem Drittanbieter zur Verfügung gestellt wurde, erstellt [52].</p>
<b>11</b>	Verblindung	Siehe 11a und 11b.
<b>11a</b>	<p>Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?</p>	<p>Bei der Studie UX023-CL303 handelt es sich um eine doppelblinde Studie (24 Wochen). Sowohl die Studienteilnehmer, als auch der Sponsor, sowie die Personen, die die Behandlung durchführten und die Zielgrößen beurteilten, waren gegenüber der Zuteilung in die Behandlungsgruppen und Verabreichung der Intervention verblindet.</p> <p>Zur Verblindung wurde den Studienteilnehmern eine Nummer zugeteilt. Weder dem Studienpersonal noch den Personen vor Ort war die Zuteilung der Nummern (Randomisierungscode) sowie die</p>

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Behandlung der Studienteilnehmer bekannt. Die Verblindung konnte nur im Falle von Sicherheitsaspekten aufgehoben werden (Entblindung Prüfarzt). Die Prüfsubstanz wurde mit der Studiennummer und einer spezifischen Kit-Nummer versehen. Die Lieferung der Prüfsubstanzen zu den Studienzentren erfolgte ebenfalls so, dass die Verblindung aufrechterhalten werden konnte.</p> <p>Die Behandlungszuteilung sollte nur in einem Notfall aufgedeckt/entblindet werden. Ein Notfall lag z. B. dann vor, wenn Sicherheitsbedenken bei der Behandlung eines Patienten bestanden oder eine Entblindung für das klinische Management oder das Wohlergehen des Patienten dringend notwendig war. Im Falle einer Entblindung musste das Datum und der Grund für die Entblindung des Patienten genau dokumentiert werden. Es wurde jedoch lediglich der Prüfarzt entblindet.</p> <p>Nach der 24-wöchigen, doppelblinden Phase der Studie bis einschließlich Woche 96 erhielten alle Studienteilnehmer Burosumab in der offenen Behandlungsphase. Sowohl die Patienten als auch die Prüfarzte blieben jedoch bis zur Analyse in Woche 48 gegenüber der ursprünglichen Behandlung verblindet. An der Extensionsphase II (Woche 97 bis ≤ 149) nahmen nur Studienteilnehmer aus den USA teil.</p>
<b>11b</b>	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	<p>Placebo setzte sich aus den gleichen Substanzen wie Burosumab zusammen, lediglich die aktive Prüfsubstanz fehlte. In beiden Fällen handelte es sich um eine durchsichtige Lösung, die in 5 ml Durchstechflaschen zur Verfügung gestellt wurde. Beide Substanzen waren mit der Studiennummer und einer spezifischen Kitnummer versehen, um die Verblindung aufrecht zu erhalten. Optisch war anhand der Verpackung kein Unterschied zwischen beiden Substanzen feststellbar.</p>
<b>12</b>	Statistische Methoden	Siehe 12a und 12b.
<b>12a</b>	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p><u>Generalized Estimating Equations (GEE)-Modell:</u></p> <p>Um Veränderungen von Baseline zu Post-Baseline zu analysieren, wurden bei wiederholten Messungen <i>Generalized Estimating Equations (GEE)</i>-Modelle verwendet. Bei dem GEE-Modell wurden</p>

		<p>als feste Effekte die Region, die Visite, der Behandlungsarm, der tatsächliche Stratifizierungsfaktor der Randomisierung (<i>BPI Durchschnittlicher Schmerz</i> &gt; 6,0 oder <math>\leq 6,0</math>) und die Interaktion der Behandlungsgruppe mit den Behandlungsvisiten berücksichtigt sowie der Baseline-Wert des entsprechenden Endpunkts als Kovariate. Die abhängige Variable war die Änderung zu Baseline des jeweiligen Endpunkts. Lediglich bei dem Endpunkt <i>BPI Schlimmster Schmerz</i> wurde der Faktor <i>BPI Durchschnittlicher Schmerz</i> nicht als Faktor in das GEE-Modell aufgenommen, da der <i>BPI Schlimmste Schmerz</i> eine hohe Korrelation mit dem <i>BPI Durchschnittlichen Schmerz</i> aufwies. Als Kovarianzstruktur wurde die <i>Compound Symmetry</i> verwendet, die für alle Studienteilnehmer im Modell die gleiche, zeitlich-konstante Korrelation annimmt.</p> <p>Die Änderungen zwischen Baseline und Post-Baseline in den Behandlungsgruppen wurden mit Hilfe der entsprechenden Punktschätzer und 95% Konfidenzintervalle beschrieben. Für die Beurteilung der statistischen Signifikanz wurden p-Werte berechnet, für die Beurteilung der Effektgröße Hedges' <i>g</i> – falls ein signifikantes Ergebnis beobachtet wurde. Für den Vergleich der Behandlungsgruppen wurden die adjustierten mittleren Änderungen (Ls Means) berechnet und mit Hilfe von Typ-III-Tests auf statistische Signifikanz geprüft.</p> <p><u>ANCOVA-Modell:</u></p> <p>Für den Vergleich der Behandlung von Baseline zu Post-Baseline wurde bei ausgewählten stetigen Wirksamkeitsendpunkten eine Kovarianzanalyse (ANCOVA-Modell) durchgeführt, die die Behandlung und den Stratifizierungsfaktor <i>BPI Durchschnittlicher Schmerz</i> (&gt; 6,0 oder <math>\leq 6,0</math>) berücksichtigte. Der Baseline-Wert des jeweiligen Endpunkts wurde als zusätzliche Kovariable eingefügt. Für den Vergleich der Behandlungsgruppen wurden die Ls Means berechnet und mit Hilfe von Typ-III-Tests auf statistische Signifikanz geprüft. Zusätzlich wurden die 95 %-KI angegeben. Das ANCOVA-Modell wurde nur bei Wirksamkeitsendpunkten angewendet, die sowohl zu Baseline als auch zu Woche 24 gemessen wurden.</p>
--	--	--

		<p><u>Analyse des primären Wirksamkeitsendpunkts:</u>  Der primäre Wirksamkeitsendpunkt, der Anteil an Studienteilnehmern, die in der Mitte des Dosisintervalls eine Serumphosphat-Konzentration oberhalb der unteren Grenze des Normbereichs von 2,5 mg/dl erreichten, gemittelt über die Dosiszyklen zwischen Baseline und Woche 24, wurde mithilfe des Cochran-Mantel-Haenszel-Tests ausgewertet (Signifikanzniveau <math>\alpha = 0,05</math>). In der Analyse wurden die Stratifizierungsfaktoren <i>BPI Durchschnittlicher Schmerz</i> und <i>Region</i> berücksichtigt.</p> <p><u>Multiples Testen:</u>  Um die <i>Family-Wise-Error-Rate</i> (FWER) bei einem Signifikanzniveau von 0,05 zu kontrollieren, wurden verschiedene Methoden für das multiple Testen bei den drei sekundären Wirksamkeitsendpunkten angewendet:</p> <ul style="list-style-type: none"> <li>• Schritt 1: Falls der Anteil an Studienteilnehmern der Burosumab-Gruppe, die eine <i>mittlere Serumphosphat-Konzentration</i> über der unteren Grenze des Normbereichs von 2,5 mg/dl in der Mitte des Dosisintervalls erreichten, gemittelt zwischen den Dosiszyklen zu Baseline und Post-Baseline statistisch signifikant größer als in der Placebo-Gruppe ist, wird mit dem 2. Schritt weiter gemacht, ansonsten wird das multiple Testen gestoppt</li> <li>• 2. Schritt: Für die drei sekundären Wirksamkeitsendpunkte (<i>BPI Schlimmster Schmerz</i>, <i>WOMAC Steifheit</i> und <i>WOMAC Physische Funktion</i>) wird die Veränderung von Baseline zu Woche 24 mit einem Signifikanzniveau von 0,05 getestet. Außerdem erfolgt eine Hochberg-Adjustierung für das multiple Testen Für die übrigen sekundären Wirksamkeitsendpunkte wird dagegen kein multiples Testen durchgeführt.</li> <li>• Laut dem Methodenpapier des IQWiG muss bei der Nutzenbewertung jedoch nicht für multiples Testen adjustiert werden, sodass die zuvor beschriebene Methode für die Nutzenbewertung nicht von Bedeutung ist.</li> </ul>
--	--	---

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<p>Figure 2: Flow Chart of Multiplicity Adjustment</p> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p><b>Primary Endpoint:</b> The proportion of subjects achieving mean serum phosphorus levels above the lower limit of normal at the mid-point of the dose interval, as averaged across dose cycles between baseline and week 24. <math>\alpha = 0.05</math></p> </div> <p style="text-align: center;">↓ If primary endpoint is significant</p> <div style="border: 1px solid black; padding: 5px;"> <p><b>Key Secondary Endpoints to be tested using Hochberg method:</b></p> <p>Change from baseline to week 24 in:</p> <ul style="list-style-type: none"> <li>• BPI Worst Pain</li> <li>• WOMAC stiffness</li> <li>• WOMAC physical function</li> </ul> <p><math>\alpha = 0.05</math></p> </div>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<p><b>Responderanalysen:</b></p> <ul style="list-style-type: none"> <li>• <i>Serumphosphat-Konzentration</i> Responder sind Studienteilnehmer, die in der Mitte des Dosisintervalls eine <i>Serumphosphat-Konzentration</i> oberhalb der unteren Grenze des Normbereichs von 2,5 mg/dl erreichten, gemittelt über die Dosiszyklen zwischen Baseline und Woche 24  Die Woche 24 Analyse wurde mithilfe eines logistischen Regressionsmodells berechnet. Das logistische Regressionsmodell wurde mit der Behandlungsgruppe, dem tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) und der Altersgruppe als unabhängige Variable durchgeführt. Das <i>Odds Ratio</i>, bei der die Burosumab-Behandlung mit Placebo verglichen wurde, wurde zusammen mit dem 95% Konfidenzintervall (KI) und dem p-Wert angegeben.</li> <li>• <i>BPI Schlimmster Schmerz</i> Responder sind Studienteilnehmer die zu Woche 24 im Vergleich zu Baseline eine 15 % bzw. 30 % Abnahme des Endpunkts <i>BPI Schlimmster Schmerz</i> aufweisen</li> <li>• <i>WOMAC Score Steifheit</i> Responder sind Studienteilnehmer, die zu Woche 24 im Vergleich zu Baseline eine Abnahme von <math>\geq 10</math> des Endpunkts <i>WOMAC Steifheit</i> aufweisen</li> <li>• <i>WOMAC Physische Funktion</i> Bei Respondern handelt es sich um diejenigen Studienteilnehmer, die zu Woche 24 im Vergleich zu Baseline eine Abnahme des Endpunkts</li> </ul>

		<p><i>WOMAC Physische Funktion</i> von <math>\geq 9,3</math> nu aufweisen</p> <p>Die Woche 24 Analysen wurden mithilfe eines logistischen Regressionsmodells berechnet. Das logistische Regressionsmodell wurde mit der Behandlungsgruppe, dem tatsächlichen Stratifizierungsfaktor (<i>BPI Durchschnittlicher Schmerz</i>) und der Altersgruppe als unabhängige Variable durchgeführt. Das <i>Odds Ratio</i>, bei der die Burosumab-Behandlung mit Placebo verglichen wurde, wurde zusammen mit dem 95% Konfidenzintervall (KI) und dem p-Wert angegeben.</p> <p><b><u>Subgruppenanalysen:</u></b></p> <p>Für die Wirksamkeitsendpunkte wurden folgende Subgruppen festgelegt:</p> <p><b>Präspezifiziert:</b></p> <ul style="list-style-type: none"> <li>• <i>Alter:</i> Subgruppen klassifiziert nach Alter zu Baseline <ul style="list-style-type: none"> <li>○ Alter <math>\leq 41</math> Jahre</li> <li>○ Alter <math>&gt; 41</math> Jahre</li> </ul> </li> <li>• <i>Ethnie:</i> Subgruppen klassifiziert nach Ethnie <ul style="list-style-type: none"> <li>○ Weiß</li> <li>○ Nicht weiß</li> </ul> </li> <li>• <i>Geschlecht:</i> Subgruppen klassifiziert nach Geschlecht <ul style="list-style-type: none"> <li>○ Männlich</li> <li>○ Weiblich</li> </ul> </li> <li>• <i>Herkunft:</i> Subgruppen klassifiziert nach Herkunft <ul style="list-style-type: none"> <li>○ Nordamerika/EU</li> <li>○ Japan</li> <li>○ Südkorea</li> </ul> </li> <li>• <i>Geplanter Stratifizierungsfaktor:</i> Subgruppen klassifiziert nach dem <i>BPI Schlimmsten Schmerz</i> zu Baseline <ul style="list-style-type: none"> <li>○ <i>BPI Schlimmster Schmerz</i> <math>\leq 6,0</math></li> <li>○ <i>BPI Schlimmster Schmerz</i> <math>&gt; 6,0</math></li> </ul> </li> <li>• <i>Verwendeter Stratifizierungsfaktor:</i> Subgruppen klassifiziert nach dem <i>BPI Durchschnittlichen Schmerz</i> zu Baseline</li> </ul>
--	--	---

		<ul style="list-style-type: none"> <li>○ <i>BPI Durchschnittlicher Schmerz</i> <math>\leq 6,0</math></li> <li>○ <i>BPI Durchschnittlicher Schmerz</i> <math>&gt; 6,0</math></li> <li>• <b>Schmerzmittel:</b> Subgruppe klassifiziert anhand der Verwendung von Schmerzmitteln zu Baseline <ul style="list-style-type: none"> <li>○ Schmerzmittel ja</li> <li>○ Schmerzmittel nein</li> </ul> </li> <li>• <b>Opioide:</b> Subgruppe klassifiziert anhand der Verwendung von Opioiden zu Baseline <ul style="list-style-type: none"> <li>○ Opioide ja</li> <li>○ Opioide nein</li> </ul> </li> <li><b>Post-hoc definierte Subgruppen:</b></li> <li>• <b>WOMAC Steifheit</b> Subgruppen klassifiziert anhand des Endpunkts <i>WOMAC Steifheit</i> zu Baseline <ul style="list-style-type: none"> <li>○ <i>WOMAC Steifheit</i> <math>\leq 62,5</math></li> <li>○ <i>WOMAC Steifheit</i> <math>&gt; 62,5</math></li> </ul> </li> <li>• <b>WOMAC Schmerzscore:</b> Subgruppen klassifiziert anhand des Endpunkts <i>WOMAC Schmerz</i> zu Baseline <ul style="list-style-type: none"> <li>○ <i>WOMAC Schmerz</i> <math>\leq 50,0</math></li> <li>○ <i>WOMAC Schmerz</i> <math>&gt; 50,0</math></li> </ul> </li> <li>• <b>WOMAC Physische Funktion:</b> Subgruppen klassifiziert anhand des Endpunkts <i>WOMAC Physische Funktion</i> zu Baseline <ul style="list-style-type: none"> <li>○ <i>WOMAC Physische Funktion</i> <math>\leq 47,8</math></li> <li>○ <i>WOMAC Physische Funktion</i> <math>&gt; 47,8</math></li> </ul> </li> <li>• <b>WOMAC Gesamtscore:</b> Subgruppen klassifiziert anhand des Endpunkts <i>WOMAC Gesamtscore</i> zu Baseline <ul style="list-style-type: none"> <li>○ <i>WOMAC Gesamtscore</i> <math>\leq 49,5</math></li> <li>○ <i>WOMAC Gesamtscore</i> <math>&gt; 49,5</math></li> </ul> </li> <li>• <b>Frakturen zu Baseline:</b> Subgruppe klassifiziert anhand des Vorhandenseins von Frakturen zu Baseline <ul style="list-style-type: none"> <li>○ Frakturen ja</li> </ul> </li> </ul>
--	--	---

Item <sup>a</sup>	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> <li>○ Frakturen nein</li> <li>• <i>6MWT</i>: Subgruppen klassifiziert anhand des <i>6MWT</i> <ul style="list-style-type: none"> <li>○ <math>6MWT \leq 365</math> m</li> <li>○ <math>6MWT &gt; 365</math> m</li> </ul> </li> </ul> <p><b>Sensitivitätsanalysen:</b> Aufgrund der fehlerhaften Stratifizierung wurden sowohl für den primären als auch die sekundären Wirksamkeitsendpunkte Sensitivitätsanalysen durchgeführt. Das logistische Regressionsmodell für den primären Endpunkt <i>Serumphosphat-Konzentration</i> beinhaltete die Behandlungsgruppe, den geplanten Stratifizierungsfaktor (<i>BPI Schlimmster Schmerz</i>) und die Altersgruppen als unabhängige Variablen. Bei den sekundären Endpunkten <i>WOMAC Steifheit</i> und <i>WOMAC Physische Funktion</i> wurde als Kovariate der geplante Stratifizierungsfaktor <i>BPI Schlimmster Schmerz</i> (<math>\leq 6</math> und <math>&gt; 6</math>) in das GEE-Modell mit einbezogen. Bei dem Endpunkt <i>BPI Schlimmster Schmerz</i> wurde dagegen der <i>BPI Durchschnittlicher Schmerz</i> als Kovariate mit aufgenommen.</p> <p>Zur Beurteilung der Robustheit der Ergebnisse und um für fehlende Werte zu adjustieren, wurden folgende Sensitivitätsanalysen durchgeführt:</p> <ul style="list-style-type: none"> <li>• <i>LOCF (Last Observation Carried Forward)</i> Die letzte nicht fehlende Beobachtung (Post-Baseline) wurde auf den entsprechenden Endpunkt zur Bewertung übertragen.</li> <li>• <i>BOCF (Baseline Observation Carried Forward)</i> Der vorhandene Baseline-Wert wurde auf den entsprechenden Endpunkt zur Bewertung übertragen.</li> <li>• <i>mBOCF (modified Baseline Observation Carried Forward)</i> Bei fehlenden Werten aufgrund eines Studienabbruchs durch UE oder Tod wurde der schlechteste Wert zwischen BOCF und LOCF verwendet andernfalls wurde LOCF verwendet</li> </ul>
<b>Resultate</b>		

Item <sup>a</sup>	Charakteristikum	Studieninformation
<b>13</b>	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	134 Erwachsene mit XLH im Alter von 18 – 65 Jahren wurden in die Studie aufgenommen und 1:1 den Behandlungsgruppen (Burosumab bzw. Placebo) zugeteilt.
<b>13a</b>	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	a) Alle 134 Erwachsene wurden auf die zwei Behandlungsgruppen 1:1 aufgeteilt b) Insgesamt erhielten 68 Studienteilnehmer Burosumab und 66 Studienteilnehmer Placebo alle 4 Wochen (28 Tage) subkutan c) Alle 134 Studienteilnehmer wurden für die Analyse des primären Zielkriteriums berücksichtigt
<b>13b</b>	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Ein Studienteilnehmer der Burosumab-Gruppe verließ die Studie 4 Wochen nach der letzten Gabe von Burosumab bevor die Untersuchungen zu Woche 24 abgeschlossen waren.
<b>14</b>	Aufnahme / Rekrutierung	Siehe untenstehendes Flussdiagramm
<b>14a</b>	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Siehe untenstehendes Flussdiagramm
<b>14b</b>	Informationen, warum die Studie endete oder beendet wurde	Das Ende der Studie, 06.12.2018, wurde als der Tag definiert, an dem die letzte Protokoll-spezifische Untersuchung (einschließlich Evaluation durch Telefonanrufe) durchgeführt wurde.
<b>a: nach CONSORT 2010.</b>		

*Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.*

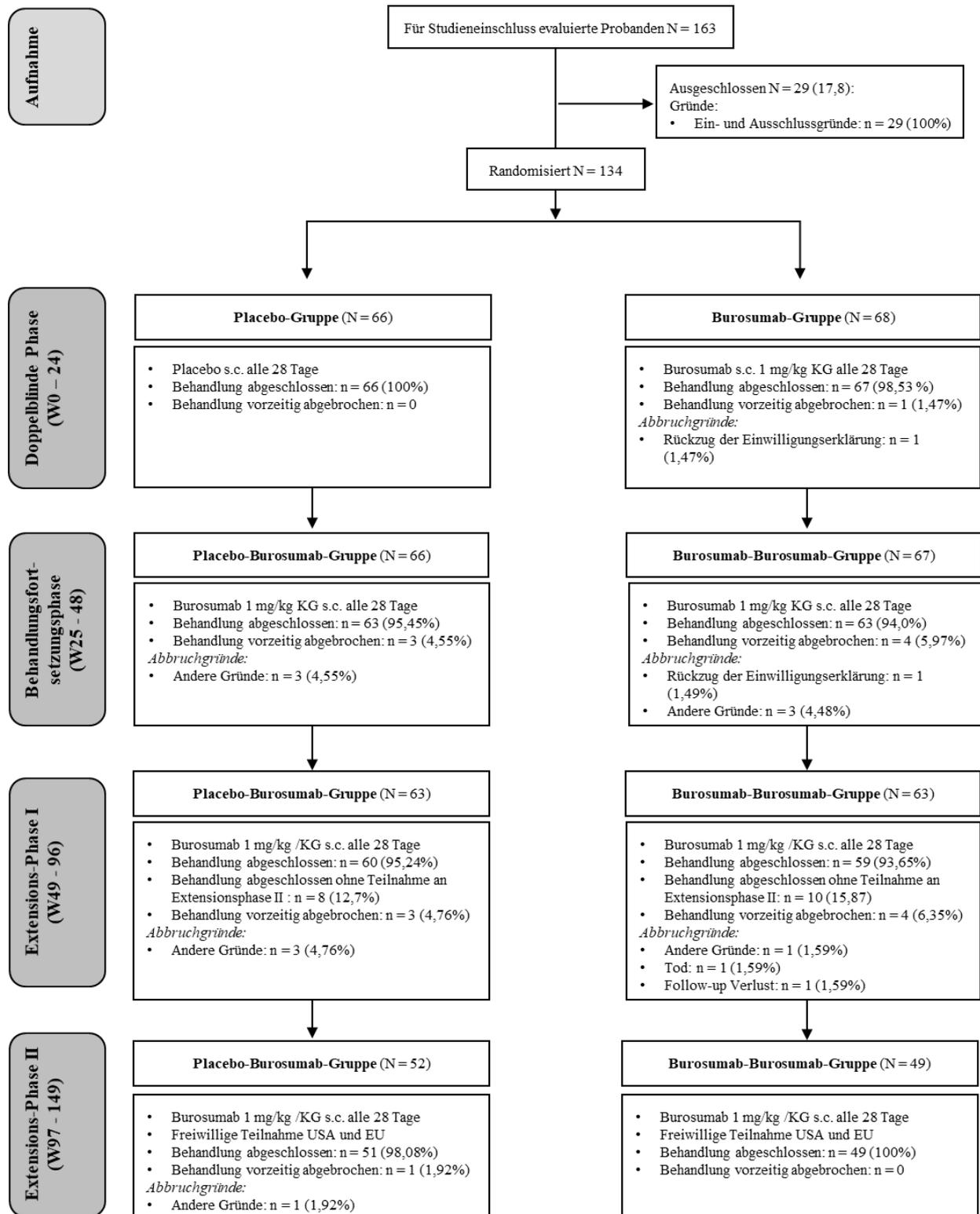


Abbildung 9: Patientenfluss der Studie UX023-CL303 gemäß CONSORT

**Anhang 4-F: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten**

Der nachfolgend dargestellte Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- Verzerrungsaspekte auf Studienebene. In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- Verzerrungsaspekte auf Endpunktebene. In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich sollen die Bögen studienbezogen ausgefüllt werden. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen ist die Blankoversion des Bogens. Dieser Blankobogen ist für jede Studie heranzuziehen. Im Anschluss daran ist ein Bewertungsbogen inklusive Ausfüllhinweisen abgebildet, der als Ausfüllhilfe dient, aber nicht als Vorlage verwendet werden soll.

*Beschreiben Sie nachfolgend die Verzerrungsaspekte jeder eingeschlossenen Studie (einschließlich der Beschreibung für jeden berücksichtigten Endpunkt). Erstellen Sie hierfür je Studie eine separate Version des nachfolgend dargestellten Bewertungsbogens.*

Tabelle 4-95 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie UX023-CL303

**Studie: UX023-CL303**

Tabelle 4-96: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
UX023-CL303 Protocol Week 24	
UX023-CL303 Statistical Analysis Plan	UX023-CL303 SAP
UX023-CL303 Clinical Study Report	UX023-CL303 CSR

#### A Verzerrungsaspekte auf Studienebene:

##### Einstufung als randomisierte Studie

**ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien

**nein** → Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

---

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde Placebo-kontrollierte Studie. Während die eine Behandlungsgruppe alle 4 Wochen Burosumab s.c. erhielt, wurde der anderen Behandlungsgruppe in dem gleichen Rhythmus s.c. Placebo verabreicht.

---

1.

##### **für randomisierte Studien: Adäquate Erzeugung der Randomisierungssequenz**

**ja**       **unklar**       **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Die Randomisierung der Studienteilnehmer erfolgte 1:1 in die beiden Behandlungsgruppen (Burosumab und Placebo) und wurde anhand des *Interactive Web Response System* (IWRS) durchgeführt. Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 Brief Pain Inventory (BPI) *Schlimmster Schmerz* des patientenberichteten Fragebogens BPI und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens BPI bestimmt. Die verkürzte Form des Fragebogens enthielt Frage 3-6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-97). Das für die Stratifizierung verwendete IWRS war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz* ≤ 6,0 oder > 6,0) verwendet wurde (Anhang 4-E 8b).

Tabelle 4-97: Nummerierung der einzelnen Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	Brief Pain Inventory
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopfschmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	- <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	- <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerz-linderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerzlinderung zeigt.
9 A – G	- <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:

Da die Studienteilnehmerrekrutierung bereits stattfand als der Fehler im IWRS entdeckt wurde, wurde beschlossen mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI schlimmster Schmerz*  $\leq 6,0$  und *BPI schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Um die Robustheit der Ergebnisse zu überprüfen, wurden zusätzlich Sensitivitätsanalysen für den primären Endpunkt *Serumphosphat* (mg/dl) und die drei sekundären Endpunkte *BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion* durchgeführt 4.2.5.4. Diese Analysen waren den primären Analysen sehr ähnlich und bestätigen somit, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Analysen in der Studie hat.

**für nicht randomisierte Studien: Zeitliche Parallelität der Gruppen**

ja     unklar     nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---



---

2.

**für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“)** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Die Randomisierung erfolgte anhand des *Interactive Web Response Systems* mit einer verdeckten Gruppenzuteilung. Weder dem Sponsor, den Prüfarzten, dem Studienpersonal noch den Studienteilnehmern war während der 24-wöchigen verblindeten Studienphase die Gruppenzuteilung bekannt.

---

**für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:3. **Verblindung von Patienten und behandelnden Personen****Patient:** ja       unklar       neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

Zur Durchführung der Studie UX023-CL303 wurden die Studienteilnehmer verblindet, sodass ihnen nicht bekannt war, ob ihnen als Prüfsubstanz Burosumab oder Placebo verabreicht wurde.

---

**behandelnde bzw. weiterbehandelnde Personen:** ja       unklar       neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

Sowohl dem Prüfarzt, dem Studienpersonal als auch dem Sponsor war die Gruppenzuteilung während der 24-wöchigen doppelblinden Phase der Studie nicht bekannt.

---

4. **Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde, Placebo-kontrollierte multizentrische Studie der Phase III. Eine Behandlungsgruppe erhielt alle 4 Wochen Burosumab und die andere Placebo s.c.. Die Stratifizierung erfolgte anhand des *BPI Durchschnittlichen Schmerzes* (*BPI Durchschnittlicher Schmerz* > 6,0 oder ≤ 6,0) ermittelt mit einem Patiententagebuch in einem Zeitraum von 7 Tagen vor Studienbeginn, sowie anhand der Region (Nordamerika/EU, Japan, Südkorea). Die Randomisierung erfolgte 1:1 in die beiden Behandlungsgruppen mittels dem *Interactive Web Response System* (IWRS). Sowohl die Patienten, als auch die Prüfarzte, das Studienpersonal und der Sponsor waren gegenüber der Behandlung verblindet. Aufgrund dessen liegt sowohl für die objektiv

---

messbaren Endpunkte, als auch für die patientenberichteten Endpunkte ein niedriges Verzerrungspotenzial vor.

#### 5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können

ja  nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Obwohl eine fehlerhafte Stratifizierung zu Beginn der Studie erfolgte, liegt dadurch keine weitere endpunktübergreifende Verzerrung vor. Dies wird im Folgenden begründet:

Die Randomisierung der Studienteilnehmer erfolgte 1:1 in die beiden Behandlungsgruppen (Burosumab und Placebo) und wurde anhand des *Interactive Web Response System* (IWRS) durchgeführt. Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 *Brief Pain Inventory* (BPI) *Schlimmster Schmerz* des patientenberichteten Fragebogens BPI und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens BPI bestimmt. Die verkürzte Form des Fragebogens enthielt Frage 3-6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3 (Tabelle 4-98). Das für die Stratifizierung verwendete IWRS war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  oder  $> 6,0$ ) verwendet wurde (Anhang 4-E 8b).

Tabelle 4-98: Nummerierung der Items des BPI-Fragebogens für die Stratifizierung in der Studie UX023-CL303

Ursprüngliche Nummerierung	Nummerierung im IWRS <sup>a</sup>	<i>Brief Pain Inventory</i>
1	- <sup>b</sup>	Die meisten von uns haben von Zeit zu Zeit Schmerzen (z. B. Kopf-schmerzen, Zahnschmerzen, bei Verstauchungen). Hatten Sie heute andere als diese Alltagsschmerzen?
2	- <sup>b</sup>	Schraffieren Sie in nachstehender Zeichnung die Gebiete, in denen Sie Schmerzen haben. Markieren Sie mit "X" die Stelle, die Sie am meisten schmerzt.
3	1	Kreisen Sie die Zahl ein, die Ihre <b>stärksten Schmerzen</b> in den letzten 24 Stunden beschreibt.
4	2	Kreisen Sie die Zahl ein, die Ihre geringsten Schmerzen in den letzten 24 Stunden beschreibt:
5	3	Kreisen Sie die Zahl ein, die Ihre <b>durchschnittlichen Schmerzen</b> in den letzten 24 Stunden beschreibt.
6	4	Kreisen Sie die Zahl ein, die aussagt, welche Schmerzen Sie in diesem Moment haben.
7	- <sup>b</sup>	Welche Behandlungen oder Medikamente erhalten Sie gegen Ihre Schmerzen?
8	- <sup>b</sup>	Bitte denken Sie an die vergangenen 24 Stunden. Wieviel Schmerzlinderung haben Sie durch Behandlungen oder Medikamente erfahren? Bitte kreisen Sie die Prozentzahl ein, die am besten die Schmerz-linderung zeigt.

9 A – G	- <sup>b</sup>	Bitte kreisen Sie die eine Zahl ein, die angibt, wie stark Ihre Schmerzen Sie in den vergangenen 24 Stunden beeinträchtigt haben:
---------	----------------	---

Da die Studienteilnehmerrekrutierung bereits stattfand, als der Fehler im IWRS entdeckt wurde, wurde beschlossen mit der Stratifizierung fortzufahren. Dies wurde jedoch als Protokollabweichung im *Electronic Data Capture System* (EDC) dokumentiert. 72,4 % und 27,6 % der Studienteilnehmer wurden anhand der Stratifizierung in die Kategorien *BPI Durchschnittlicher Schmerz*  $\leq 6,0$  bzw. *BPI Durchschnittlicher Schmerz*  $> 6,0$  eingeteilt. Bezogen auf den ursprünglich geplanten Stratifizierungsfaktor entsprachen 32,8 % bzw. 67,2 % der Studienteilnehmer den Kategorien *BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ . Bei 60 % der Studienteilnehmer entsprach der durchschnittliche Schmerz (*BPI durchschnittlicher Schmerz*  $\leq 6,0$  und *BPI Durchschnittlicher Schmerz*  $> 6,0$ ) dem schlimmsten Schmerz (*BPI Schlimmster Schmerz*  $\leq 6,0$  und *BPI Schlimmster Schmerz*  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Um die Robustheit der Ergebnisse zu überprüfen, wurden zusätzlich Sensitivitätsanalysen für den primären Endpunkt *Serumphosphat* (mg/dl) und die drei sekundären Endpunkte *BPI Schlimmster Schmerz*, *WOMAC Steifheit* und *WOMAC Physische Funktion* durchgeführt 4.2.5.4. Diese Analysen waren den primären Analysen sehr ähnlich und belegen, dass die fehlerhafte Stratifizierung keinen Einfluss auf die Analysen in der Studie hat.

**Dies bestätigt somit, dass in der Studie UX023-CL303 keine endpunktübergreifende Verzerrung vorliegt.**

#### Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

niedrig       hoch

Begründung für die Einstufung:

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde, Placebo-kontrollierte multizentrische Studie der Phase III, die erwachsene XLH-Patienten im Alter von 18-65 Jahren beinhaltet. Die Randomisierung der Studienteilnehmer erfolgte 1:1 in die beiden Behandlungsgruppen (Burosumab und Placebo) und wurde anhand des *Interactive Web Response System* (IWRS) durchgeführt. Ursprünglich sollte die Stratifizierung in der Studie UX023-CL303 anhand der Frage 3 *Brief Pain Inventory (BPI) Schlimmster Schmerz* des patientenberichteten Fragebogens BPI und der Region erfolgen. Hierfür wurde der Schmerz anhand eines Patiententagebuchs 7 Tage vor der Baseline-Visite sowie am Tag der Visite mit einer verkürzten Form des Fragebogens BPI bestimmt. Die verkürzte Form des Fragebogens enthielt Frage 3-6 der ursprünglichen Version. Da Frage 1 (Vorhandensein von Schmerzen, die sich von Alltagsschmerzen unterscheiden) und Frage 2 (Lokalisation der Schmerzen) nicht mit aufgeführt wurden, erfolgte eine neue Nummerierung der einzelnen Items, sodass Frage 3 (*BPI Schlimmster Schmerz*) als Frage 1 klassifiziert wurde und Frage 5 (*BPI Durchschnittlicher Schmerz*) als Frage 3. Das für die Stratifizierung verwendete IWRS war jedoch nicht daran angepasst, sodass für die Stratifizierung nicht Frage 1 (*BPI Schlimmster Schmerz*) sondern Frage 3 (*BPI Durchschnittlicher Schmerz*  $\leq 6,0$  oder  $> 6,0$ ) verwendet wurde (Anhang 4-E 8b).

Da die Rekrutierung und Verabreichung der Prüfsubstanz bereits begonnen hatte als der Fehler entdeckt wurde, wurde beschlossen, die Stratifizierung weiterhin anhand des *BPI Durchschnittlichen Schmerzes* ( $\leq 6,0$  oder  $> 6,0$ ) durchzuführen. Der Stratifizierungsfehler wurde als Protokollabweichung im elektronischen Datenerfassungssystem (EDC) aufgezeichnet. Analysen zeigten, dass 60 % der Studienteilnehmer, unabhängig davon ob es sich um den *BPI Durchschnittlichen Schmerz* oder *BPI Schlimmsten Schmerz* handelt, der gleichen Stratifizierungskategorie zugeordnet worden wären ( $\leq 6,0$  oder  $> 6,0$ ). Der Pearson Korrelationskoeffizient  $r$  zwischen den Gruppen betrug 0,82256 ( $p < 0,001$ ) und stützt die Annahme, dass der Einfluss der fehlerhaften Stratifizierung minimal ist. Des Weiteren erfolgten Sensitivitätsanalysen des primären und der drei sekundären Wirksamkeitsendpunkte um die Robustheit der Ergebnisse zu überprüfen, die ebenfalls bestätigten, dass der Fehler in der Stratifizierung zu keiner Verzerrung der Ergebnisse führte. Sowohl die Studienteilnehmer als auch das Studienpersonal, die Prüfarzte und der Sponsor waren während der 24-wöchigen doppelblinden Phase

der Studie gegenüber der Behandlung verblindet. **Aufgrund dessen ist bei der vorliegenden Studie von einem niedrigen Verzerrungspotenzial auszugehen.**

---

**B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:****Endpunkt: Serumphosphat-Konzentration****1. Verblindung der Endpunkterheber**

ja     unklar     nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Die Messung der *Serumphosphat-Konzentration* erfolgte in einem zentralen Labor. Das Studienpersonal und Laborpersonal war gegenüber der Behandlung sowie der Laborwerte, deren Veränderungen mit einer Burosumab-Behandlung assoziiert werden konnten, verblindet.

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja     unklar     nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um die Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja     unklar     nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja     nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig     hoch

Begründung für die Einstufung:

---

Bei der Studie UX023-CL303 handelt es sich um eine randomisierte, doppelblinde Placebo-kontrollierte multizentrische Studie der Phase III. Die durchgeführten Analysen beinhalteten die ITT-Population und erfolgten nach dem Prinzip der ergebnisunabhängigen Berichterstattung. Der objektive numerische Endpunkt *Serumphosphat* weist daher ein **niedriges Verzerrungspotenzial** auf.

---

**Endpunkt: Brief Pain Inventory****1. Verblindung der Endpunkterheber**

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Aufgrund des randomisierten, doppelblinden Studiendesigns waren die Prüfarzte, die Studienteilnehmer und das Studienpersonal gegenüber der Zuordnung und der Behandlung der Studienteilnehmer verblindet.

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja       unklar       nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um diejenigen Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja       unklar       nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja       nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig       hoch

Begründung für die Einstufung:

---

Bei dem *BPI*-Fragebogen handelt es sich um einen patientenberichteten Schmerz-spezifischen Fragebogen, in dem die Schmerzintensität abgefragt wird (*BPI Schlimmster Schmerz*, *BPI Geringster Schmerz*, *BPI Durchschnittlicher Schmerz*). Ursprünglich wurde dieser Fragebogen für Krebspatienten entwickelt, ist jedoch mittlerweile auch bei verschiedenen anderen Erkrankungen wie z. B. bei Patienten mit Arthrose oder Kreuzschmerzen etabliert und stellt ein valides Messinstrument zur Erfassung der Schmerzen bei Patienten mit XLH dar. Der Endpunkt *BPI Schlimmster Schmerz* bezieht sich auf den schlimmsten Schmerz innerhalb der letzten 24 Stunden und wird anhand einer Skala von 0-10 ermittelt. Eine ergebnisunabhängige Berichterstattung sowie die Umsetzung des ITT-Prinzips waren während der gesamten Studiendauer gewährleistet. Aufgrund der patientenberichteten Form des Fragebogens und

Verblindung des Studienpersonals, der Studienärzte und Studienteilnehmer, liegt somit ein **niedriges Verzerrungspotenzial** des Endpunkts *Brief Pain Inventory (BPI Schlimmster Schmerz, BPI Schmerzintensität und BPI Schmerzinterferenz)* vor.

---

**Endpunkt: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)**

**1. Verblindung der Endpunkterheber**

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Sowohl die Studienteilnehmer, das Studienpersonal, der Sponsor als auch der Prüfarzt waren während der gesamten Studiendauer von 24 Wochen bzgl. der Behandlung der Studienteilnehmer verblindet.

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja       unklar       nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um diejenigen Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja       unklar       nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja       nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

niedrig       hoch

Begründung für die Einstufung:

Der WOMAC Fragebogen ist ein patientenberichteter Fragebogen, der die 2 Domänen Physische Funktion (17 Fragen) und WOMAC Steifheit (2 Fragen) beinhaltet und sich auf die Einschätzung des Patienten innerhalb der letzten 48 Stunden bezieht. Die Einstufung erfolgt anhand einer 5-Punkte-Likert-Skala, bei der ein höherer Score eine stärkere Beeinträchtigung indiziert. Etabliert wurde der Fragebogen ursprünglich für Patienten mit Hüft- und Kniearthrose, ist aber auch mittlerweile für XLH validiert [36]. Die Inhaltsvalidität ist anhand der Guidelines von der FDA überprüft und gewährleistet. Während der gesamten Studiendauer erfolgte eine ergebnisunabhängige Berichterstattung unter Einbezug der ITT-Population. Aufgrund der patientenberichteten Form des Fragebogens und Verblindung des Studienpersonals, der Studienärzte und Studienteilnehmer, liegt somit ein **niedriges Verzerrungspotenzial** bei dem Endpunkt **WOMAC (WOMAC Steifheit und WOMAC Physische Funktion)** vor.

---

**Endpunkt: Brief Fatigue Inventory (BFI)****1. Verblindung der Endpunkterheber**

ja     unklar     nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Aufgrund des randomisierten, doppelblinden Studiendesigns waren die Prüfarzte, die Studienteilnehmer und das Studienpersonal gegenüber der Zuordnung und der Behandlung der Studienteilnehmer verblindet.

---

**2. Adäquate Umsetzung des ITT-Prinzips**

ja     unklar     nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um diejenigen Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

ja     unklar     nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

ja     nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

**niedrig**                       **hoch**

Begründung für die Einstufung:

---

Der Fragebogen *Brief Fatigue Inventory* ist ein patientenberichteter Fragebogen, der insgesamt 9 Punkte zur Fatigue beinhaltet und sich auf die Einschätzung des Patienten innerhalb der letzten 24 Stunden bezieht. Die Einstufung erfolgt anhand einer Skala von 0-10, bei der ein höherer Score eine stärkere Beeinträchtigung indiziert. Etabliert wurde der Fragebogen ursprünglich für Krebspatienten, um den Einfluss der Erkrankung und Therapie auf das Level der Fatigue zu evaluieren. Zwischenzeitlich wurde die Validität des Fragebogens auch für Patienten mit Erkrankungen des Muskel-Skelettsystems wie z. B. Rheumatoider Arthritis und Arthrose. Während der gesamten Studiendauer erfolgte eine ergebnisunabhängige Berichterstattung unter Einbezug der ITT-Population. Aufgrund der patientenberichteten Form des Fragebogens und Verblindung des Studienpersonals, der Studienärzte und Studienteilnehmer, liegt somit ein **niedriges Verzerrungspotenzial** bei dem Endpunkt *Brief Fatigue Inventory (BFI Schlimmste Fatigue und BFI Globale Fatigue)* vor.

---

**Endpunkt: 6-Minuten-Gehdistanz Test (6MWT)**

**1. Verblindung der Endpunkterheber**

**ja**                       **unklar**                       **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Aufgrund des randomisierten, doppelblinden Studiendesigns waren die Prüfer, die Studienteilnehmer und das Studienpersonal gegenüber der Zuordnung und der Behandlung der Studienteilnehmer verblindet.

---

**2. Adäquate Umsetzung des ITT-Prinzips**

**ja**                       **unklar**                       **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um diejenigen Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

**ja**                       **unklar**                       **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.

---

**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja       neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.

---

**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig       hoch

Begründung für die Einstufung:

---

Der *6-Minuten-Gehdistanz Test (6MWT)*, ist ein objektives Messinstrument zur Evaluation der körperlichen Funktion. Die Patienten mussten innerhalb von 6 Minuten eine vorabgemessene Wegstrecke zurücklegen. Anschließend wurde die zurückgelegte Gesamtdistanz gemessen. Hierfür wurden die Kriterien der *American Thoracic Society Guidelines* als qualitätssichernde Maßnahme berücksichtigt. Die Durchführung erfolgte durch vorab geschulte und lizenzierte Physiotherapeuten. Während der gesamten Studiendauer fand eine ergebnisunabhängige Berichterstattung unter Einbezug der ITT-Population statt. Aufgrund der patientenberichteten Form des Fragebogens und Verblindung des Studienpersonals, der Studienärzte und Studienteilnehmer, liegt somit ein **niedriges Verzerrungspotenzial** bei dem Endpunkt *6MWT* vor.

---

**Endpunkt: Patient Global Impression of Improvement (PGI)****1. Verblindung der Endpunkterheber** ja       unklar       neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

---

Aufgrund des randomisierten, doppelblinden Studiendesigns waren die Prüfarzte, die Studienteilnehmer und das Studienpersonal gegenüber der Zuordnung und der Behandlung der Studienteilnehmer verblindet.

---

**2. Adäquate Umsetzung des ITT-Prinzips** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um diejenigen Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.

---

#### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja       nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.

---

#### Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig       hoch

Begründung für die Einstufung:

---

Der patientenberichtete Fragebogen *Patient Global Impression of Improvement (PGI-I)* ist ein weit verbreitetes Messinstrument in klinischen Studien, um die patientenindividuelle Wahrnehmung der Wirksamkeit einer Behandlung auf chronische Schmerzen zu evaluieren. Die wahrgenommene Verbesserung bzw. Verschlechterung wird anhand einer 7-Punkte Skala, die von sehr stark verbessert bis sehr stark verschlechtert reicht, analysiert. Während der gesamten Studiendauer fand eine ergebnisunabhängige Berichterstattung unter Einbezug der ITT-Population statt. Aufgrund der patientenberichteten Form des Fragebogens und Verblindung des Studienpersonals, der Studienärzte und Studienteilnehmer, liegt somit ein **niedriges Verzerrungspotenzial** bei dem Endpunkt *PGI-I* vor.

---

#### Endpunkt: Sicherheit

##### 1. Verblindung der Endpunkterheber

ja       unklar       nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

Aufgrund des randomisierten, doppelblinden Studiendesigns waren die Prüfer, die Studienteilnehmer und das Studienpersonal gegenüber der Zuordnung und der Behandlung der Studienteilnehmer verblindet.

---

##### 2. Adäquate Umsetzung des ITT-Prinzips

ja       unklar       nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Alle Patienten des primären Analysesets wurden für die Auswertungen herangezogen. Hierbei handelt es sich um diejenigen Studienteilnehmer, die mindestens eine Dosis der Prüfsubstanz während der Placebo-kontrollierten Phase erhielten.

---

**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja       unklar       neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

Aufgrund der Verblindung ist eine ergebnisunabhängige Berichterstattung gewährleistet.**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja       neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

Keine sonstigen (endpunktübergreifenden) Aspekte vorhanden.**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig       hoch

Begründung für die Einstufung:

---

Es handelt sich um eine randomisierte, doppelblinde Placebo-kontrollierte multizentrische Phase III-Zulassungsstudie. Sowohl das ITT-Prinzip als auch eine ergebnisunabhängige Berichterstattung wurden umgesetzt. Es sind keine sonstigen (endpunktspezifischen) Aspekte, die zur Verzerrung der Endpunktergebnisse führen können, gegeben. Entsprechend den Vorgaben des G-BA werden die Nebenwirkungen über die UEs und SUEs sowohl als Gesamtraten als auch aufgeteilt nach SOC und PT dargestellt. Es erfolgte eine Einstufung des kausalen Zusammenhangs der UEs mit der Prüfmedikation sowie des Schweregrads der UEs gemäß der *National Cancer Institute Common Terminology Criteria for Adverse Events* (NST CTCAE) durch den verantwortlichen Prüfarzt. Aufgrund dessen kann von einem **niedrigen Verzerrungspotenzial** für den Endpunkt *Sicherheit* ausgegangen werden.

---

Hinweis: Der nachfolgend dargestellte Bewertungsbogen mit Ausfüllhinweisen dient nur als Ausfüllhilfe für den Blankobogen. Er soll nicht als Vorlage verwendet werden.

### Bewertungsbogen zur Beschreibung von Verzerrungsaspekten (Ausfüllhilfe)

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunkübergreifend; B: endpunktspezifisch).

#### A Verzerrungsaspekte auf Studienebene:

##### Einstufung als randomisierte Studie

**ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien

**nein:** Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anzeichen vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

---

---

1.

##### für randomisierte Studien:

##### Adäquate Erzeugung der Randomisierungssequenz

**ja:** Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

**unklar:** Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

**nein:** Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

##### für nicht randomisierte Studien:

##### Zeitliche Parallelität der Gruppen

**ja:** Die Gruppen wurden zeitlich parallel verfolgt.

**unklar:** Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

**nein:** Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

2.

**für randomisierte Studien:****Verdeckung der Gruppenzuteilung („allocation concealment“)** **ja:** Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

 **unklar:** Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau. **nein:** Die Gruppenzuteilung erfolgte nicht verdeckt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

**für nicht randomisierte Studien:****Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** **ja:** Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anzeichen dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

 **unklar:** Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau. **nein:** Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

**3. Verblindung von Patienten und behandelnden Personen****Patient:** **ja:** Die Patienten waren verblindet. **unklar:** Es finden sich keine diesbezüglichen Angaben. **nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.Angaben zum Kriterium; obligate Begründung für die Einstufung:

**behandelnde bzw. weiterbehandelnde Personen:**

**ja:** Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, die primär behandelnde Person (z. B. Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

**unklar:** Es finden sich keine diesbezüglichen Angaben.

**nein:** Aus den Angaben geht hervor, dass die behandelnden Personen nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

**4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte**

*Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).*

*Beispiele zu a und b:*

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

*Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter*

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

*Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:*

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

*Zulässige Gründe sind:*

- *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
- *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
- *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anzeichen für eine ergebnisgesteuerte Darstellung eines Endpunkts zu Verzerrungen der Ergebnisse der übrigen Endpunkte führen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anzeichen dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, sind Verzerrungen für die anderen Endpunkte möglich. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Verzerrung der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen. Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und dies nicht zur Verzerrung anderer Endpunkte führt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anzeichen für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---



---

##### 5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrung führen können

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien
- intransparenter Patientenfluss
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:
  - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
  - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein.
  - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.
  - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.

- ja**
- nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---



---

##### Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

- niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

- hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:

---



---

### B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

**Endpunkt:** \_\_\_\_\_

#### 1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---



---

#### 2. Adäquate Umsetzung des ITT-Prinzips

Kommen in einer Studie Patienten vor, die die Studie entweder vorzeitig abgebrochen haben oder wegen Protokollverletzung ganz oder teilweise aus der Analyse ausgeschlossen wurden, so sind diese ausreichend genau zu beschreiben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen zu berücksichtigen (i. d. R. ITT-Analyse, siehe Äquivalenzstudien). Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

**unklar:** Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

**nein:** Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

### 3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

*Beachte die Hinweise zu Punkt A.4!*

**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

**nein:** Es liegen Anzeichen für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

---

### 4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

*z. B.*

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

**ja**

**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

---

---

### **Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):**

*Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.*

**niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

**hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:

---

---

### **Anhang 4-G: Ergänzende Analysen**

Ergänzende *post-hoc* durchgeführte Analysen zu den Abschnitten 4.3.1.3.1 und 4.3.1.3.2 sind separat im Dokument 2022-01-31\_Modul4B\_Burosumab\_Anhang4-G.pdf aufgeführt.