

# Nutzenbewertung



Gemeinsamer  
Bundesausschuss

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

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

**Wirkstoff: Maralixibat**

Datum der Veröffentlichung: 17. April 2023

## Inhaltsverzeichnis

|   |    |
|---|----|
| Tabellenverzeichnis .....   | 3  |
| Abbildungsverzeichnis .....   | 5  |
| Abkürzungsverzeichnis .....   | 6  |
| Hintergrund .....   | 7  |
| 1 Fragestellung .....   | 8  |
| 2 Auswahl und Methodik der relevanten Studien .....                           | 9  |
| 2.1 Studienbasis für die Nutzenbewertung .....                                | 9  |
| 2.2 Beschreibung der eingeschlossenen Studie .....                            | 10 |
| 2.3 Endpunkte .....   | 18 |
| 2.3.1 Mortalität .....  | 18 |
| 2.3.2 Morbidität .....  | 19 |
| 2.3.3 Lebensqualität .....  | 29 |
| 2.3.4 Sicherheit .....  | 30 |
| 2.3.5 Übersicht der Erhebungszeitpunkte .....                                 | 33 |
| 2.4 Statistische Methoden .....   | 34 |
| 2.5 Verzerrungspotential auf Studien- und Endpunktebene .....                 | 38 |
| 2.6 Indirekter Vergleich .....  | 40 |
| 2.6.1 Zusammenfassende Bewertung der Methodik des indirekten Vergleichs ..... | 43 |
| 3 Ergebnisse der eingeschlossenen Studien .....                               | 46 |
| 3.1 Studiencharakteristika und Studienmedikation .....                        | 46 |
| 3.2 Mortalität .....  | 52 |
| 3.3 Morbidität .....  | 52 |
| 3.4 Lebensqualität .....  | 62 |
| 3.5 Sicherheit .....  | 64 |
| 4 Diskussion der Methodik und Ergebnisse .....                                | 69 |
| 4.1 Zulassungsstatus und Zulassungspopulation von Maralixibat .....           | 69 |
| 4.2 Design und Methodik der Studien .....                                     | 70 |
| 4.3 Mortalität .....  | 73 |
| 4.4 Morbidität .....  | 73 |
| 4.5 Lebensqualität .....  | 77 |
| 4.6 Sicherheit .....  | 77 |
| 5 Anforderung an eine qualitätsgesicherte Anwendung .....                     | 79 |
| 6 Zusammenfassung der Nutzenbewertung .....                                   | 80 |
| Referenzen .....  | 86 |
| Anhang .....  | 89 |

## Tabellenverzeichnis

|   |    |
|---|----|
| Tabelle 1: Übersicht über die Studienbasis .....  | 9  |
| Tabelle 2: Charakterisierung der Studie ICONIC .....  | 10 |
| Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie ICONIC.....   | 13 |
| Tabelle 4: Charakterisierung der Intervention in der Studie ICONIC .....  | 14 |
| Tabelle 5: Charakterisierung der Studie MRX-801 .....   | 15 |
| Tabelle 6: Charakterisierung der Intervention in der Studie MRX-801 .....   | 17 |
| Tabelle 7: Zusammenfassung der Endpunktbewertung der Studien ICONIC und MRX-801 .....   | 18 |
| Tabelle 8: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie ICONIC .....  | 33 |
| Tabelle 9: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie MRX-801.....  | 34 |
| Tabelle 10: Verzerrungspotential der RW-Phase der Studie ICONIC.....  | 38 |
| Tabelle 11: Verzerrungspotential der in der Nutzenbewertung berücksichtigten Endpunkte der RW-Phase der Studie ICONIC .....   | 39 |
| Tabelle 12: Schritt 1 der Selektion von Kontrollen: Identifikation der Personen und Visiten innerhalb der GALA-Studie .....   | 42 |
| Tabelle 13: Allgemeine Angaben; Studien ICONIC und MRX-801.....   | 46 |
| Tabelle 14: Charakterisierung der Studienpopulationen zu Baseline der Studien ICONIC und Studie MRX-801 .....   | 48 |
| Tabelle 15: Begleitmedikationen zur Behandlung des Pruritus bei mehr als 10 % der Studienpopulation nach Substanz; Studie ICONIC.....                                 | 51 |
| Tabelle 16: Anteil der Tage bzw. Personen mit einem Wert von $\leq 1$ Punkt im ItchRO in der Erhebung am Morgen; Studie ICONIC, RW-Phase (Woche 19 bis Woche 22)..... | 53 |
| Tabelle 17: Veränderung des Pruritus im ItchRO(Obs) in der RW-Phase; Studie ICONIC.....   | 54 |
| Tabelle 18: Veränderung des Pruritus im ItchRO(Pt) in der RW-Phase; Studie ICONIC, RW-Population $\geq 5$ Jahre zu Baseline.....                                      | 55 |
| Tabelle 19: Responderanalysen zur Verbesserung im ItchRO in der Erhebung am Morgen; Studie ICONIC, ITT-Population.....  | 57 |
| Tabelle 20: Veränderung des Pruritus im ItchRO, Wert der Erhebung am Morgen; Studie ICONIC, ITT-Population.....   | 58 |
| Tabelle 21: Klinische Kratzskala, Responderanalyse; Studie ICONIC, ITT-Population .....   | 58 |
| Tabelle 22: Veränderung anthropometrischer Parameter; Studie ICONIC, ITT-Population.....  | 59 |
| Tabelle 23: Veränderung in der PedsQL-Fatigue (Parent) in der RW-Phase; Studie ICONIC, RW-Population $\geq 2$ Jahre .....   | 61 |
| Tabelle 24: Veränderung in der PedsQL-Fatigue (Child) in der RW-Phase; Studie ICONIC, RW-Population ab 5 Jahre .....  | 61 |
| Tabelle 25: Veränderung in der PedsQL-Fatigue; Studie ICONIC, ITT-Population ab 2 Jahren (Parent) und ab 5 Jahren (Child).....  | 62 |

|  |    |
|--|----|
| Tabelle 26: Veränderung im PedsQL (Parent) in der RW-Phase;<br>Studie ICONIC, RW-Population (Woche 18 bis Woche 22) .....  | 63 |
| Tabelle 27: Veränderung im PedsQL (Child) in der RW-Phase;<br>Studie ICONIC, RW-Population ab 5 Jahre (Woche 18 bis Woche 22) .....  | 63 |
| Tabelle 28: Veränderung der Lebensqualität im PedsQL; Studie ICONIC, ITT-Population .....  | 64 |
| Tabelle 29: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE<br>während der Behandlungsphase; Studien ICONIC und MRX-801 .....   | 65 |
| Tabelle 30: UE mit Inzidenz $\geq 10$ % nach MedDRA-Systemorganklassen und Preferred Terms;<br>Studien ICONIC und MRX-801.....   | 65 |
| Tabelle 31: Schwere UE (Grad 3 oder höher), die bei $\geq 5$ % der Personen aufgetreten sind,<br>nach Systemorganklasse und Preferred Term; Studien ICONIC und MRX-801 ..... | 66 |
| Tabelle 32: SUE, die bei $\geq 5$ % der Personen aufgetreten sind, nach Systemorganklasse und<br>Preferred Term; Studien ICONIC und MRX-801, Sicherheitspopulation .....     | 67 |
| Tabelle 33: UE von besonderem Interesse; Studie ICONIC .....   | 68 |
| Tabelle 34: Zusammenfassende Darstellung der vergleichenden Ergebnisse<br>der Studie ICONIC – RW-Phase .....   | 81 |
| Tabelle 35: Zusammenfassende Darstellung der Ergebnisse<br>der Studie ICONIC – Unkontrollierte Maralixibat-Behandlung .....  | 83 |
| Tabelle 36: Zusammenfassende Darstellung der Ergebnisse der Studie MRX-801 .....   | 85 |
| Tabelle 37: Diagnostische Kriterien des ALGS .....   | 89 |
| Tabelle 38: Veränderung des sBA-Werts in der RW-Phase; Studie ICONIC, mITT-Population.....   | 90 |
| Tabelle 39: Anteil der Tage bzw. Personen mit einem Wert $\leq 1$ Punkt<br>im ItchRO in der Erhebung am Morgen; Studie ICONIC.....   | 90 |

## Abbildungsverzeichnis

|   |    |
|---|----|
| Abbildung 1: Schematische Darstellung des Studienablaufs der Studie ICONIC [5] .....  | 13 |
| Abbildung 2: Übersicht über die Datenquellen der gepoolten Maralixibat-Kohorte für den indirekten Vergleich .....   | 41 |
| Abbildung 3: Mittelwert der Änderung des wöchentlichen morgendlichen ItchRO(Obs) gegenüber Baseline nach Behandlungsgruppe bis Woche 48 in der Studie ICONIC..... | 56 |
| Abbildung 4: Mittelwert der Änderung des wöchentlichen morgendlichen ItchRO(Pt) gegenüber Baseline nach Behandlungsgruppe bis Woche 48 in der Studie ICONIC.....  | 57 |

## Abkürzungsverzeichnis

|                |  |
|----------------|--|
| ALGS           | Alagille-Syndrom   |
| ALT            | Alanin-Aminotransferase  |
| AM-NutzenV     | Arzneimittel-Nutzenbewertungsverordnung                              |
| ANCOVA         | Analysis of Covariance   |
| CIC            | Caregiver Impression of Change                                       |
| CTCAE          | Common Terminology Criteria for Adverse Events                       |
| EPAR           | European Public Assessment Report                                    |
| FDA            | U.S. Food and Drug Administration                                    |
| G-BA           | Gemeinsamer Bundesausschuss  |
| GALA           | Global ALagille Alliance   |
| GGT            | Gamma-Glutamyltransferase  |
| GKV            | Gesetzliche Krankenversicherung                                      |
| IQWiG          | Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen     |
| ItchRO         | Itch Reported Outcome  |
| ItchRO(Obs)    | Itch Reported Outcome (Observer)                                     |
| ItchRO(Pt)     | Itch Reported Outcome (Patient)                                      |
| ITT            | Intention-to-Treat   |
| KI             | Konfidenzintervall   |
| LOCF           | Last Observation Carried Forward                                     |
| LS             | Least Squares  |
| MedDRA         | Medical Dictionary for Regulatory Activities                         |
| mITT           | modifizierte Intention-to-Treat                                      |
| MMRM           | Mixed Model Repeated Measures  |
| N              | Anzahl   |
| PedsQL         | Pediatric Quality of Life Inventory                                  |
| PedsQL-Fatigue | Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale |
| PFIC           | Progressive familiäre intrahepatische Cholestase                     |
| PIC            | Patient Impression of Change   |
| PT             | Preferred Term   |
| pU             | pharmazeutischer Unternehmer   |
| SAP            | Statistischer Analyseplan  |
| sBA            | Serumgallensäure   |
| SGB            | Sozialgesetzbuch   |
| SOC            | Systemorganklasse  |
| SUE            | Schwerwiegendes unerwünschtes Ereignis                               |
| UE             | Unerwünschtes Ereignis   |
| VerfO          | Verfahrensordnung des G-BA   |
| RW             | Randomisierte Absetzphase (Randomized Withdrawal)                    |

## Hintergrund

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

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

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

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

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Maralixibat in seiner Sitzung am 13. April 2023 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 16. Januar 2023 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pU in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 17. April 2023 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

## 1 Fragestellung

Maralixibat (Livmarli®) ist ein Arzneimittel mit dem Status „Orphan Drug“. Gemäß § 35a Absatz 1 Satz 11 1. Halbs. SGB V gilt für Orphan Drugs der medizinische Zusatznutzen bereits durch die Zulassung als belegt. Der G-BA bestimmt bei Orphan Drugs, die einen Umsatz von 30 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. In die Bewertung werden gemäß Zulassungsstatus Personen in folgender Indikation eingeschlossen [6]:

Alagille-Syndrom (ALGS) ab dem Alter von 2 Monaten zur Behandlung des cholestatischen Pruritus.

Gemäß Fachinformation beträgt die empfohlene Zieldosis 380 µg/kg einmal täglich. Die Anfangsdosis beträgt 190 µg/kg einmal täglich und sollte nach einer Woche auf 380 µg/kg einmal täglich erhöht werden. Dabei entspricht Maralixibat 380 µg/kg/Tag als freie Base einer Dosis von Maralixibat-Chlorid von 400 µg/kg/Tag.

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

| Studien-<br>bezeichnung                 | Studie mit<br>Dossier<br>vom pU<br>eingereicht | Studie vom<br>pU als rele-<br>vant für die<br>Ableitung<br>des Zusatz-<br>nutzens<br>erachtet | Studie<br>relevant für<br>die Nutzen-<br>bewertung | Ausschlussgrund   |
|---|--|---|--|-------------------|
| <b>Studien zum Wirkstoff</b>            |  |   |  |                   |
| LUM001-304<br>(ICONIC) <sup>1)</sup>    | Ja   | Ja  | Ja   | -                 |
| MRX-801 <sup>1)</sup>                   | Ja   | Nein  | Ja   | -                 |
| <b>Studien zum indirekten Vergleich</b> |  |   |  |                   |
| GALA-MRX-ALGS <sup>2)</sup>             | Ja   | Ja  | Nein   | Siehe Kapitel 2.6 |

<sup>1)</sup> Zulassungsrelevante Studie gemäß EPAR [5].

<sup>2)</sup> Im Rahmen des Zulassungsverfahrens bei der EMA ergänzend zu den Zulassungsstudien berücksichtigt. In den indirekten Vergleich gehen Personen folgender Studien mit dem Wirkstoff ein: LUM001-304 (ICONIC), LUM001-301/303 und LUM001-302/305. Die Kontrollkohorte besteht aus Personen der GALA-Studie.

Abkürzungen: EMA: European Medicines Agency; EPAR: European Public Assessment Report; GALA: Global ALagille Alliance; pU: pharmazeutischer Unternehmer.

#### Zur Nutzenbewertung für Maralixibat herangezogene Studien und Daten

- Herstellerdossier zu Maralixibat [22]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [5]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der Studie LUM001-304 (ICONIC) [24,25,26,27] und der Studie MRX-801 [19,20]
- Publikation der Studie ICONIC von Gonzales et al. [7]

## 2.2 Beschreibung der eingeschlossenen Studie

Die Unterlagen zur Nutzenbewertung für Maralixibat basieren auf der Zulassungsstudie LUM001-304 (ICONIC) und der Studie MRX-801. Die Studien und die Intervention werden in den Tabellen 2 bis 6 charakterisiert.

### Studie ICONIC

Tabelle 2: Charakterisierung der Studie ICONIC

| Charakteristikum                | Beschreibung  |
|---------------------------------|---|
| <b>Design und Studienablauf</b> | <p>ICONIC ist eine offene Langzeitstudie mit einer doppelblinden, Placebo-kontrollierten, randomisierten Absetzphase bei Kindern mit ALGS, zur Untersuchung der Sicherheit und Wirksamkeit von Maralixibat (MRX).</p> <p>Die Studie ist in 6 Teile gegliedert:</p> <ul style="list-style-type: none"> <li>• Offene Dosisescalationsphase über 6 Wochen mit Dosierungen von bis zu 400 µg/kg/Tag<sup>1</sup>.</li> <li>• 12-wöchige, offene, Behandlungsphase mit stabiler Dosierung.</li> <li>• RW-Phase: 4-wöchige randomisierte, doppelblinde, Placebo-kontrollierte Absetzphase (stratifiziert nach Ansprechen im sBA-Wert (definiert als Reduktion von ≥ 50 % zu Woche 12 oder 18): ja; nein).</li> <li>• 26-wöchige, langfristige Behandlungsphase mit stabilen Dosierungen bis zu 400 µg/kg/Tag.</li> <li>• Optionale Behandlungsverlängerung über 52 Wochen.</li> <li>• Optionale langfristige Behandlungsverlängerung, wobei MRX auf maximal 800 µg/kg/Tag (400 µg/kg zweimal täglich) erhöht werden kann, je nach Wirksamkeit und Sicherheit.</li> </ul> <p>Die langfristige Behandlungsverlängerung dauert an, bis einer der folgenden Fälle eintritt: i) Probanden kommen für die Teilnahme an einer anderen Maralixibat-Studie in Frage; ii) MRX ist im Handel erhältlich; (iii) der Sponsor stellt das Programm oder die Entwicklung in dieser Indikation ein.</p> <p>Für die optionale Behandlungsverlängerung über 52 Wochen<sup>2</sup> und die optionale langfristige Behandlungsverlängerung<sup>3</sup> bestehen wiederum Ein- und Ausschlusskriterien.</p> <p>Studienteilnehmende, die Protokolländerung 4 (Implementierung der optionalen langfristigen Behandlungsverlängerung) nicht zustimmen, setzen ihre Behandlung fort, wie in Protokolländerung 3 (optionale Behandlungsverlängerung) beschrieben.</p> |
| <b>Population</b>               | <p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Personen im Alter von 12 Monaten bis einschließlich 18 Jahren.</li> <li>• ALGS-Diagnose basierend auf definierten diagnostischen Kriterien (siehe Anhang).</li> <li>• Nachweis einer Cholestase (durch eines oder mehrerer der folgenden Kriterien): <ul style="list-style-type: none"> <li>○ Gesamt-sBA &gt; 3 × ULN für das betreffende Alter.</li> <li>○ Konjugiertes Bilirubin &gt; 1 mg/dl.</li> <li>○ Mangel an fettlöslichen Vitaminen, der nicht durch andere Ursachen erklärbar ist.</li> <li>○ GGT &gt; 3x ULN für das betreffende Alter.</li> <li>○ Hartnäckiger Juckreiz (Pruritus), der nur durch eine Lebererkrankung erklärbar ist.</li> </ul> </li> <li>• Durchschnittlicher Tageswert &gt; 2 im ItchRO-Tagebuch (maximal möglicher Tageswert = 4) für 2 aufeinanderfolgende Wochen im Screening-Zeitraum vor der Dosierung. Tageswert ist der maximale Wert aus den Erhebungen am Morgen und am Abend; durchschnittlicher Tageswert ist die Summe aller Tageswerte dividiert durch die Anzahl der Tage, für die der ItchRO ausgefüllt wurde. Basis für den Einschluss waren die Betreuer-Einschätzungen (ItchRO(Obs)).</li> </ul>   |

| Charakteristikum   | Beschreibung   |
|--|--|
|  | <ul style="list-style-type: none"> <li>• Mindestens 10 Erhebungen im ItchRO (morgens oder abends) während 2 aufeinanderfolgender Wochen des Screening-Zeitraums (maximal mögliche Anzahl der Erhebungen = 14 pro Woche).</li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Chronische Diarrhö, die eine ständige intravenöse Flüssigkeits- oder Nahrungszufuhr erfordert.</li> <li>• Chirurgische Unterbrechung des enterohepatischen Kreislaufs.</li> <li>• Vorherige Lebertransplantation.</li> <li>• Dekompensierte Leberzirrhose, Vorgeschichte oder Vorhandensein von klinisch signifikantem Aszites, Varizenblutung und/oder Enzephalopathie.</li> <li>• Vorgeschichte oder Vorhandensein einer anderen begleitenden Lebererkrankung.</li> <li>• Vorgeschichte oder Vorhandensein einer anderen Krankheit oder eines Zustands, der bekanntermaßen die Absorption, Verteilung, den Stoffwechsel oder die Ausscheidung von Arzneimitteln beeinflusst, einschließlich Gallensalz-Metabolismus im Darm (z. B. entzündliche Darmerkrankungen).</li> <li>• Vorgeschichte oder Vorhandensein von Gallen- oder Nierensteinen.</li> <li>• Bekannte HIV-Infektion.</li> <li>• Krebserkrankungen, mit Ausnahme von In-situ-Karzinomen oder Krebserkrankungen, die mind. 5 Jahre vor Screening ohne Anzeichen eines Wiederauftretens behandelt wurden.</li> <li>• Verabreichung von Gallensäure- oder Lipid-bindenden Granulaten innerhalb von 28 Tagen vor Screening und während der gesamten Studie.</li> <li>• Gewicht von mehr als 50 kg bei Screening.</li> </ul> |
| <b>Intervention und Zahl der Personen;<br/>Datenschnitte</b> | <p>Gescreent gesamt: N = 36<br/> Eingeschlossen: N = 31<br/> Intervention: N = 31<sup>4)</sup></p> <p>RW-Phase:<br/> Kontrolle: N = 16<sup>4)</sup><br/> Intervention: N = 13</p> <p>Interim-Datenschnitt: 01.12.2019<br/> Interim-Studienbericht: 13.10.2020<br/> Abschluss der Studie: 28.05.2020<br/> Finaler Datenschnitt: 21.08.2020<br/> Datum finaler Studienbericht: 11.11.2020</p>  |
| <b>Ort und Zeitraum der Durchführung</b>                     | <p><b>Studienzentren</b><br/> 9 Studienzentren in Frankreich (3), Australien (2), Belgien, Polen, Spanien, Großbritannien (jeweils 1).</p> <p><b>Studienzeitraum</b><br/> Die Patientinnen und Patienten wurden zwischen dem 28.10.2014 und 11.09.2015 in die Studie aufgenommen. Studienende: 28.05.2020.</p>   |

| Charakteristikum                        | Beschreibung  |
|---|---|
| <b>Endpunkte gemäß Studienprotokoll</b> | <p><b>Primärer Endpunkt</b><br/>Veränderung des Nüchtern-sBA-Levels zu Woche 22 gegenüber Woche 18 bei Personen, die zuvor auf die MRX-Behandlung angesprochen haben.</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• ALP</li> <li>• ALT</li> <li>• Gesamtbilirubin</li> <li>• Direktes Bilirubin</li> <li>• Pruritus gemessen mittels ItchRO(Obs) und ItchRO(Pt)</li> <li>• Veränderung des Nüchtern-sBA-Levels (von Baseline zu Woche 18)</li> </ul> <p><b>Weitere Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Pruritus-Responder gemessen mittels ItchRO</li> <li>• Responder in der klinischen Kratzskala</li> <li>• Nüchtern-sBA-Level</li> <li>• Andere biochemische Marker der Cholestase (Gesamtcholesterin, LDL-C)</li> <li>• Gallensäure-Synthese (Serum 7<math>\alpha</math>-Hydroxy-4-cholesten-3-on [7<math>\alpha</math>C4])</li> <li>• PedsQL</li> <li>• PedsQL-Fatigue</li> <li>• PedsQL – Family Impact Module</li> <li>• PIC</li> <li>• CIC</li> <li>• Caregiver Global Therapeutic Benefit</li> <li>• Veränderung der Xanthome</li> <li>• Veränderung der Körpergröße und des Körpergewichts</li> <li>• Schmackhaftigkeit der MRX-Formulierung über die Zeit</li> <li>• Plasmalevel von MRX</li> </ul> <p><b>Sicherheit</b></p> <ul style="list-style-type: none"> <li>• Unerwünschte Ereignisse</li> <li>• Laborparameter</li> <li>• Alpha-Fetoprotein-Werte</li> <li>• Befunde körperlicher Untersuchungen und Vitalzeichen</li> </ul> |

<sup>1)</sup> Die eingesetzte Dosis von 400  $\mu\text{g}/\text{kg}/\text{Tag}$  MRX-Chlorid entspricht einer Dosis von 380  $\mu\text{g}/\text{kg}/\text{Tag}$  MRX als freie Base.

<sup>2)</sup> Relevante Ein-/Ausschlusskriterien für die 52-wöchige optionale Behandlungsverlängerung gemäß Protokoll-amendment 3: Beendigung des Protokolls bis Woche 48 ohne Sicherheitsprobleme. Kein chirurgischer Eingriff zur Unterbrechung des enterohepatischen Kreislaufs.

<sup>3)</sup> Relevante Ein-/Ausschlusskriterien für die optionale langfristige Behandlungsverlängerung gemäß Protokoll-amendment 4: Beendigung des Protokolls bis Woche 48 ohne Sicherheitsprobleme bzw. eventuelle Sicherheitsprobleme standen nicht mit MRX in Zusammenhang, Laborergebnisse kehren auf akzeptables Niveau zurück und kein Kriterium für einen Studienabbruch war erfüllt. Sofern eine Unterbrechung der MRX-Einnahme von  $\geq 7$  Tagen vorlag, bestanden die gleichen Ausschlusskriterien wie zu Studienbeginn, ausgenommen der Gewichtsbeschränkung auf 50 kg.

<sup>4)</sup> Aufgrund des Studiendesigns erhielten alle eingeschlossenen Personen MRX. In der RW-Phase wurden 16 Personen mit einem Placebo behandelt.

Abkürzungen: ALGS: Alagille-Syndrom; ALP: Alkalische Phosphatase; ALT: Alanin-Aminotransferase; CIC: Caregiver Impression of Change; GGT: Gamma-Glutamyltransferase; ItchRO: Itch Reported Outcome; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); LDL-C: Low-Density-Lipoprotein-Cholesterol; MRX: Maralixibat; PedsQL: Pediatric Quality of Life Inventory; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; PIC: Patient Impression of Change; RW: Randomisierte Absetzphase; sBA: Serumgallensäure; ULN: Upper Limit of Normal.

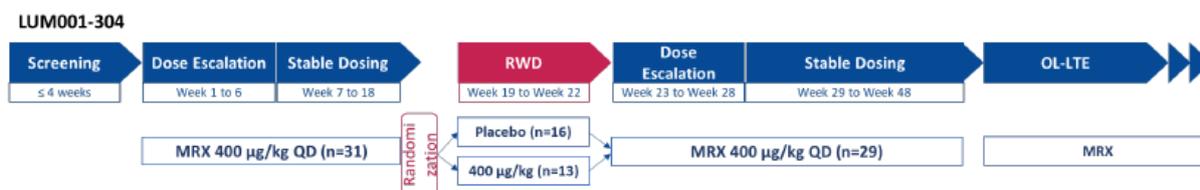


Abbildung 1: Schematische Darstellung des Studienablaufs der Studie ICONIC [5]

### Protokolländerungen

Es wurden 6 Änderungen des Originalprotokolls vom 20.03.2014 vorgenommen. Alle Änderungen erfolgten nach Einschluss der ersten Person am 28.10.2014. Die letzte Person wurde im September 2015 in die Studie eingeschlossen, von daher erfolgten die Amendments 3–5.1 nach Einschluss aller Patientinnen und Patienten. Unklar ist, wie viele Personen sich zu Amendment 1 und 2 jeweils bereits in der Studie befanden.

Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie ICONIC

| Amendment                   | Wesentliche Änderungen   |
|-----------------------------|--|
| Amendment 1<br>06.03.2015   | Vorhandensein oder die Vorgeschichte von Gallen- oder Nierensteinen und bekannter Überempfindlichkeit gegen MRX oder einen seiner Bestandteile als zusätzliche Ausschlusskriterien hinzugefügt.  |
| Amendment 2<br>08.05.2015   | <ul style="list-style-type: none"> <li>• Gewicht von mehr als 50 kg bei Screening als Ausschlusskriterium hinzugefügt; in der Folge Reduktion der Maximaldosis von 30 auf 20 mg/Tag.</li> <li>• Korrektur der Responder-Definition im sBA-Wert für die mITT-Population von Woche 18 auf Woche 12.</li> <li>• Beschreibung einer geplanten Erweiterungsstudie.</li> </ul>   |
| Amendment 3<br>13.11.2015   | <ul style="list-style-type: none"> <li>• Aufnahme einer optionalen Behandlungsverlängerung (nach Woche 48) inklusive Ein- und Ausschlusskriterien.</li> <li>• Generierung von Daten zu Sicherheit und Wirksamkeit bei Personen, die langfristig mit MRX behandelt werden, einschließlich Genotypisierungsmerkmale.</li> <li>• Symptome der Erkrankung sollen nicht weiter als UE klassifiziert werden, sofern sie im Rahmen normaler täglicher Schwankung oder Bestandteil der erwarteten Progression sind, die bereits bei der Wirksamkeit abgebildet ist.</li> </ul>   |
| Amendment 4<br>28.03.2017   | <ul style="list-style-type: none"> <li>• Hinzufügen einer optionalen langfristigen Behandlungsverlängerung, sodass die Behandlung solange fortgesetzt werden kann, bis definierte Ereignisse eintreten, inklusive Ein- und Ausschlusskriterien.</li> <li>• Klarstellung, dass geeignete Personen, die zuvor aus der Studie ausgeschieden waren, wieder in die Studie eintreten und in der optionalen Follow-up-Behandlungsphase eine Studienbehandlung erhalten können.</li> <li>• Korrektur der Responder-Definition im sBA-Wert für die mITT-Population von Woche 12 auf Woche 12 oder 18.</li> <li>• Anpassung des Stratifizierungskriteriums der Randomisierung, Responder im sBA zu Woche 12, auf Woche 12 oder 18.</li> <li>• Bestimmung, dass die Dosierung während der Studie angepasst werden kann, wenn sich das Gewicht der Testperson um <math>\geq 10\%</math> verändert.</li> <li>• Aufnahme einer unverblindeten Interimsanalyse nachdem alle Testpersonen Woche 48 erreicht haben oder die Studie abgebrochen haben.</li> <li>• Festlegung, dass dieselbe Version des PedsQL während der gesamten Studie eingesetzt wird, unabhängig von Geburtstagen während der Studie.</li> </ul> |
| Amendment 5<br>06.11.2017   | Entblindung der Studie, da alle Studienteilnehmenden Woche 48 erreicht oder die Studie abgebrochen haben.  |
| Amendment 5.1<br>08.01.2019 | Keine relevanten Änderungen (Wechsel des Sponsors der Studie).   |

Abkürzungen: mITT: modifizierte Intention-to-Treat; MRX: Maralixibat; PedsQL: Pediatric Quality of Life Inventory; sBA: Serumgallensäure; UE: Unerwünschtes Ereignis.

## Charakterisierung der Intervention

*Tabelle 4: Charakterisierung der Intervention in der Studie ICONIC*

| <b>Intervention</b>  | <b>Kontrolle</b>   |
|--|--|
| <b>Offene Behandlungsphase (Woche 1–18)</b>  |  |
| <p><b>Maralixibat</b><br/> Einmal täglich oral am Morgen in einer Lösung von 1,0 ml (Gewicht <math>\geq</math> 10 kg) oder 0,5 ml (Gewicht <math>&lt;</math> 10 kg).<br/> Offene Dosisescalationsperiode: Wöchentliche Erhöhung über einen Zeitraum von 6 Wochen auf 400 <math>\mu</math>g/kg/Tag oder ein Maximum von 20 mg/Tag wie folgt:<br/> Woche 1: 14 <math>\mu</math>g/kg/Tag MRX QD<br/> Woche 2: 35 <math>\mu</math>g/kg/Tag MRX QD<br/> Woche 3: 70 <math>\mu</math>g/kg/Tag MRX QD<br/> Woche 4: 140 <math>\mu</math>g/kg/Tag MRX QD<br/> Woche 5: 280 <math>\mu</math>g/kg/Tag MRX QD<br/> Woche 6: 400 <math>\mu</math>g/kg/Tag MRX QD (bei maximaler täglicher Dosis von 20 mg QD)<br/> Woche 7–18: Offene Behandlungsphase bei stabiler Dosierung (Woche 6 Dosis)</p> <p>Im Fall von Unverträglichkeiten aufgrund gastrointestinaler Symptome (z. B. Durchfall, Magen-Darm-Symptome wie Durchfall, Bauchschmerzen, Krämpfe) kann eine Dosisreduktion erfolgen. Eine spätere Eskalation der Dosis ist untersagt.</p>  |  |
| <b>RW-Phase (Woche 19–22)</b>  |  |
| <p><b>Maralixibat</b><br/> 400 <math>\mu</math>g/kg/Tag MRX QD oral oder die höchste tolerierte Dosis unter 400 <math>\mu</math>g/kg/Tag.</p>  | <p><b>Placebo</b><br/> Orale Lösung von 1,0, 0,5 oder 0,25 ml ohne MRX.</p>  |
| <b>Langfristige Behandlungsphase (Woche 23–48)<sup>1)</sup></b>  |  |
| <p><b>Maralixibat</b></p> <ul style="list-style-type: none"> <li>• Weiterhin 400 <math>\mu</math>g/kg/Tag MRX QD oder die höchste tolerierte Dosis unter 400 <math>\mu</math>g/kg/Tag.</li> <li>• Simulierte Dosisescalation zur Aufrechterhaltung der Verblindung.</li> </ul>   | <p><b>Maralixibat</b></p> <ul style="list-style-type: none"> <li>• Dosisescalation über 6 Wochen wie zu Therapiebeginn (s. o.) auf 400 <math>\mu</math>g/kg/Tag oder ein Maximum von 20 mg/Tag.</li> <li>• Stabile Dosierung für weitere 20 Wochen.</li> </ul> |
| <b>Optionale Behandlungsverlängerung (52 Wochen)<sup>2)</sup></b>  |  |
| <ul style="list-style-type: none"> <li>• MRX-Dosisunterbrechung <math>&lt;</math> 7 Tage: Fortsetzung von MRX in der Dosis zu Woche 48.</li> <li>• MRX-Dosisunterbrechung <math>\geq</math> 7 Tage: Dosisescalation von 35, über 70 (Woche 49), 140, 280 auf 400 <math>\mu</math>g/kg/Tag, falls die Dosisstufen jeweils toleriert wurden.</li> </ul>  |  |
| <b>Optionale langfristige-Behandlungsverlängerung<sup>3)</sup></b>   |  |
| <ul style="list-style-type: none"> <li>• Maralixibat Dosisunterbrechung nach 52-wöchiger Behandlungserweiterung <math>&lt;</math> 7 Tage: Evaluierung für Dosisescalation in Form einer zusätzlichen Verabreichung am Nachmittag. <ul style="list-style-type: none"> <li>○ Keine Dosisescalation: Fortsetzung der bislang verabreichten Dosierung.</li> <li>○ Dosisescalation<sup>4)</sup>: Nachmittagsdosisescalation beginnend mit 140 <math>\mu</math>g/kg für 4 Wochen gefolgt von 400 <math>\mu</math>g/kg oder der maximal verträglichen Dosis MRX.</li> </ul> </li> <li>• Maralixibat Dosisunterbrechung nach 52-wöchiger Behandlungserweiterung <math>\geq</math> 7 Tage: Zunächst Dosisescalation innerhalb von 5 Wochen von 35 auf bis zu 400 <math>\mu</math>g/kg/Tag. Gefolgt von Evaluierung für Dosisescalation in Form einer zusätzlichen Verabreichung am Nachmittag. <ul style="list-style-type: none"> <li>○ Keine Dosisescalation: Fortsetzung der bislang verabreichten Dosierung.</li> <li>○ Dosisescalation<sup>4)</sup>: Nachmittagsdosisescalation beginnend mit 140 <math>\mu</math>g/kg für 4 Wochen gefolgt von 400 <math>\mu</math>g/kg oder der maximal verträglichen Dosis MRX.</li> </ul> </li> </ul> |  |

| Intervention  | Kontrolle |
|---|-----------|
| <p><b>Nicht erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Gallensäure-bindende Harze.</li> <li>• Neu initiierte Medikationen zur Behandlung des Pruritus in den ersten 22 Behandlungswochen.</li> </ul> <p><b>Erlaubte Begleitmedikation</b></p> <p>Begleitmedikationen in einer stabil gehaltenen Dosis in den ersten 22 Behandlungswochen, nicht weiter spezifiziert.</p> |           |

- <sup>1)</sup> In der langfristigen Behandlungsphase erfolgen Dosisanpassungen bei Bedarf aufgrund einer Änderung des Körpergewichts der Patientinnen oder Patienten um  $\geq 10\%$  gegenüber Screening.
- <sup>2)</sup> Nur Patientinnen und Patienten, die nach definierten Kriterien geeignet waren.
- <sup>3)</sup> Nur Patientinnen und Patienten, die nach definierten Kriterien geeignet waren; Behandlungsende nicht fest definiert, bis zum Eintreten eines der folgenden Ereignisse: Eignung der/des Studienteilnehmenden für eine andere MRX-Studie; kommerzielle Verfügbarkeit; Beendigung des Studienprogramms durch den Sponsor.
- <sup>4)</sup> Nur bei stabiler Morgendosis für mind. 4 Wochen. Die Maximaldosis ist  $800\ \mu\text{g}/\text{kg}/\text{Tag}$ , maximal jedoch  $50\ \text{mg}/\text{Tag}$ .
- Abkürzungen: MRX: Maralixibat; QD: einmal täglich; RW: Randomisierte Absetzphase.

## Studie MRX-801

Tabelle 5: Charakterisierung der Studie MRX-801

| Charakteristikum                | Beschreibung  |
|---------------------------------|---|
| <b>Design und Studienablauf</b> | <p>MRX-801 ist eine offene, multizentrische Phase-II-Studie zur Untersuchung der Sicherheit und Verträglichkeit von Maralixibat (MRX) bei der Behandlung von Säuglingen (&lt; 12 Monate) mit cholestatischen Lebererkrankungen (PFIC oder ALGS).</p> <p>Die Studie ist in folgende Abschnitte unterteilt:</p> <ul style="list-style-type: none"> <li>• Screening (bis zu 4 Wochen)</li> <li>• Dosisescalation (Woche 1–6)</li> <li>• Erhaltungsphase (Woche 7–13)</li> <li>• Langzeitphase (mind. bis zu einem Alter von 12 Monaten)</li> </ul> <p>Sobald ein Alter von <math>\geq 12</math> Monaten erreicht ist, verbleiben die Teilnehmenden in der Langzeitphase, bis einer der folgenden Fälle eintritt:</p> <ul style="list-style-type: none"> <li>• Berechtigung zur Teilnahme an einem erweiterten Zugangsprogramm.</li> <li>• Kommerzielle Verfügbarkeit des Medikaments.</li> <li>• Sponsor stellt das Programm ein.</li> </ul> <p>Da für die vorliegende Nutzenbewertung ausschließlich Personen mit ALGS relevant sind, wird im Folgenden nur die ALGS-Kohorte der Studie berücksichtigt und dargestellt.</p>                                     |
| <b>Population</b>               | <p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Personen im Alter von unter 12 Monaten (mind. 3 mussten unter 9 Monate sein).</li> <li>• Gestationsalter <math>\geq 36</math> Wochen bei Geburt. Für Kinder, die mit einem Gestationsalter zwischen 32 und 36 Wochen geboren werden, ist ein postmenstruelles Alter von <math>\geq 36</math> Wochen erforderlich.</li> <li>• Körpergewicht <math>\geq 2,5\ \text{kg}</math>.</li> <li>• ALGS-Diagnose basierend auf definierten diagnostischen Kriterien (siehe Anhang).</li> <li>• Nachweis einer Cholestase (durch eines oder mehrerer der folgenden Kriterien): <ul style="list-style-type: none"> <li>○ Gesamt-sBA <math>&gt; 2 \times \text{ULN}</math> für das betreffende Alter.</li> <li>○ Konjugiertes Bilirubin <math>&gt; 1\ \text{mg}/\text{dl}</math>.</li> <li>○ Mangel an fettlöslichen Vitaminen, der nicht durch andere Ursachen erklärbar ist.</li> <li>○ GGT <math>&gt; 3 \times \text{ULN}</math> für das betreffende Alter.</li> <li>○ Hartnäckiger Juckreiz (Pruritus), der nur durch eine Lebererkrankung erklärbar ist.</li> </ul> </li> </ul> |

| Charakteristikum   | Beschreibung  |
|--|---|
|  | <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Chronische Diarrhö, die eine ständige intravenöse Flüssigkeits- oder Nahrungszufuhr erfordert.</li> <li>• Chirurgische Unterbrechung des enterohepatischen Kreislaufs.</li> <li>• Vorherige Lebertransplantation oder Lebertransplantation dringend erforderlich.</li> <li>• Dekompensierte Leberzirrhose (INR &gt; 1,5 [unresponsiv zur Vitamin-K-Therapie], Albumin &lt; 3,0 g/dl, Vorgeschichte oder Vorhandensein von klinisch signifikantem Aszites, Varizenblutung und/oder Enzephalopathie).</li> <li>• ALT oder Gesamtbilirubin im Serum &gt; 15 x ULN.</li> <li>• Vorgeschichte oder Vorhandensein einer anderen begleitenden Lebererkrankung.</li> <li>• Vorgeschichte oder Vorhandensein einer anderen Krankheit oder eines Zustands, der bekanntermaßen die Absorption, Verteilung, den Stoffwechsel oder die Ausscheidung von Arzneimitteln beeinflusst, einschließlich Gallensalz-Metabolismus im Darm (z. B. entzündliche Darmerkrankungen).</li> <li>• Lebermasse in der Bildgebung mit Verdacht auf ein hepatozelluläres Karzinom.</li> </ul> |
| <b>Intervention und Zahl der Personen; Datenschnitte</b> | <p>Gescreent gesamt: N = 8<br/> Eingeschlossen: N = 8<br/> Intervention: N = 8</p> <p>Interim-Datenschnitt 04.05.2022<br/> Interim-Studienbericht: 20.07.2022<br/> Laufende Studie; Studienabschluss für August 2023 geplant<sup>1)</sup>.</p>  |
| <b>Ort und Zeitraum der Durchführung</b>                 | <p><b>Studienzentren</b><br/> 15 Studienzentren in den USA, Großbritannien, Frankreich, Belgien, Polen.</p> <p><b>Studienzeitraum</b><br/> Studienstart: 09.09.2021<sup>1)</sup><br/> Laufende Studie.</p>  |
| <b>Endpunkte gemäß Studienprotokoll</b>                  | <p><b>Primärer Endpunkt</b><br/> Sicherheit (UE) und Verträglichkeit.</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Nüchtern-sBA-Level zu Woche 13</li> <li>• Veränderung der Leberwerte ALT, AST und Bilirubin</li> <li>• Veränderung der fettlöslichen Vitamine</li> <li>• Pharmakokinetik</li> </ul> <p><b>Explorative Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Pruritus gemessen mittels ItchRO</li> <li>• Pruritus gemessen mittels klinischer Kratzskala</li> <li>• Wachstum (Größe, Gewicht, Körpermaße)</li> <li>• Nutzung des Gesundheitssystems (z. B. Hospitalisierungen)</li> <li>• Belastung der betreuenden Personen über Fehltag bei der Arbeit</li> </ul>  |

<sup>1)</sup> Gemäß Clinicaltrials.gov [21].

Abkürzungen: ALGS: Alagille-Syndrom; ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; GGT: Gamma-Glutamyltransferase; INR: International Normalized Ratio; ItchRO: Itch Reported Outcome; PFIC: Progressive familiäre intrahepatische Cholestase; sBA: Serumgallensäure; ULN: Upper Limit of Normal.

### Protokolländerungen

Insgesamt wurden 11, teils landesspezifische, Änderungen am Protokoll vorgenommen. Eine Auflistung der implementierten Veränderungen konnte den Studienunterlagen nicht entnommen werden. Da die Angabe zum Einschluss der ersten Person zudem nicht plausibel ist (09.09.2022, und damit nach Interim-Datenschnitt), ist unklar wie viele Amendments nach dem Einschluss von Testpersonen erfolgten. Erst mit Protokollamendment 4 vom 14.01.2022 wurde die Erhebung des ItchRO(Obs) aufgenommen. Es ist unklar ob für die Studienteilnehmenden überhaupt Baseline-Daten zum ItchRO vorliegen.

### Charakterisierung der Intervention

*Tabelle 6: Charakterisierung der Intervention in der Studie MRX-801*

| Intervention  |
|---|
| <p><b>Maralixibat in einer oralen Lösung (mit 5, 10, 15 und 20 mg/ml)</b><br/>           Einmal täglich, möglichst am Morgen vor einer Mahlzeit.<br/>           Dosisescalationsperiode: 200 µg/kg/Tag MRX für mind. eine Woche.<br/>           Erhaltungsperiode: 400 µg/kg/Tag MRX oder die höchste tolerierte Dosis für die restliche Studie.<br/>           Im Alter von &lt; 1 Monat beträgt die Dosis 75 µg/kg/Tag für die Dauer der Behandlung unterhalb dieser Altersgrenze.</p> <p>Im Fall von Unverträglichkeiten (z. B. gastrointestinale Symptome) kann eine Dosisreduktion erfolgen; mind. 200 µg/kg/Tag sind jedoch erforderlich, um die Studie fortzusetzen. Bei einem UE CTCAE-Grad 4, spezifizierten abnormalen Leberwerten sowie einer schweren Diarrhö, die eine Hospitalisierung und/ oder intravenöse Zuführung von Flüssigkeit oder Nährstoffen bedarf, ist eine Dosisunterbrechung erforderlich.</p> |
| <p><b>Nicht erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Propylenglykol-haltige Mittel.</li> <li>• Inhibitoren des ilealen Gallensäure-Transporters (IBATi).</li> <li>• Weitere Prüfpräparate.</li> <li>• Präparate, die durch Alkoholdehydrogenasen metabolisiert werden.</li> </ul>  |
| <p><b>Erlaubte Begleitmedikation</b><br/>           Alle Behandlungen, die zum Behandlungsstandard zählen (z. B. Supplemente fettlöslicher Vitamine).</p>   |

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; MRX: Maralixibat; UE: Unerwünschtes Ereignis.

## 2.3 Endpunkte

In diesem Kapitel wird die Eignung der Endpunkte hinsichtlich Operationalisierung, Patientenrelevanz und Validität beurteilt. Dazu wurden das Herstellerdossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt. Alle Endpunkte der eingeschlossenen Studien wurden einer Prüfung unterzogen. Endpunkte, die in der nachfolgenden Tabelle nicht gelistet sind, wurden weder vom pU noch im Rahmen der Nutzenbewertung als patientenrelevant bzw. bewertungsrelevant eingestuft. Tabelle 7 stellt das Ergebnis dieser Bewertung zusammenfassend dar.

Tabelle 7: Zusammenfassung der Endpunktbewertung der Studien ICONIC und MRX-801

| Zusammenfassung der Endpunktbewertung   | Kategorie      | Studie |           | Berücksichtigung im Dossier des pU | Berücksichtigung in der Nutzenbewertung |
|---|----------------|--------|-----------|------------------------------------|---|
|   |                | ICONIC | MRX-801   |                                    |   |
| Todesfälle  | Mortalität     | •      | •         | Ja                                 | Ja                                      |
| Pruritus mittels <ul style="list-style-type: none"> <li>• ItchRO</li> <li>• Eindruck der Veränderung (PIC/CIC)</li> <li>• Klinische Kratzskala</li> </ul> | Morbidity      | •      | •         | Ja                                 | Ja                                      |
|   |                | •      |           | Ja                                 | Nein                                    |
|   |                | •      | •         | Ja                                 | Ergänzend                               |
| sBA <sup>1)</sup>   |                | •      | •         | Ergänzend                          | Ergänzend                               |
| Körperliche Entwicklung   |                | •      | •         | Ja                                 | Ja                                      |
| Xanthomatose mittels <ul style="list-style-type: none"> <li>• Xanthom-Skala</li> <li>• Veränderung der Xanthome</li> </ul>                                |                | •      |           | Ja                                 | Nein                                    |
|   |                | •      |           | Nein                               | Nein                                    |
| Leberfunktion   | •              | •      | Ergänzend | Nein                               |   |
| PedsQL-Fatigue  | •              |        | Ja        | Ja                                 |   |
| PedsQL  | Lebensqualität | •      |           | Ja                                 | Ja                                      |
| Unerwünschte Ereignisse <sup>2)</sup>   | Sicherheit     | •      | •         | Ja                                 | Ja                                      |

<sup>1)</sup> Primärer Endpunkt der Studie ICONIC.

<sup>2)</sup> Primärer Endpunkt der Studie MRX-801.

Abkürzungen: ItchRO: Itch Reported Outcome; PedsQL: Pediatric Quality of Life Inventory; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; pU: pharmazeutischer Unternehmer; sBA: Serumgallensäure.

### 2.3.1 Mortalität

#### Todesfälle

Der Endpunkt „Todesfälle“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Todesfälle wurden sowohl in der Studie ICONIC als auch in der Studie MRX-801 im Rahmen der Sicherheitserfassung über UE, die zum Tod führen, erhoben und waren nicht als eigenständiger Endpunkt definiert.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

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

### Validität

Gemäß Protokoll der Studie ICONIC sollten nur Todesfälle aufgrund von UE erfasst werden. Laut SAP sollten jedoch alle während der Studie aufgetretenen Todesfälle dokumentiert werden. In Studie MRX-801 sollten alle Todesfälle erfasst werden.

## **2.3.2 Morbidität**

### **2.3.2.1 Pruritus**

#### **ItchRO**

Der Endpunkt „ItchRO“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Das elektronische Patiententagebuch Itch Reported Outcome (ItchRO) zur Erfassung des Pruritus kam in der Studie ICONIC sowohl in einer betreuerberichteten (ItchRO(Obs)) als auch einer patientenberichteten (ItchRO(Pt)) Version, zum Einsatz. Dabei wurde der ItchRO(Pt) von allen Patientinnen und Patienten ab 9 Jahren beantwortet oder bereits ab 5 Jahren, gegebenenfalls unter Hilfestellung durch die betreuende Person. Beide Versionen wurden zweimal täglich erhoben, am Morgen und am Abend. Im Folgenden sind mit der Bezeichnung „ItchRO“ immer beide Versionen des Instruments gemeint.

Im Fall des ItchRO(Pt) wurden die Patientinnen und Patienten täglich am Morgen gefragt, wie stark es sie in der Nacht, bis zum Aufwachen, juckte. Die Antwortoptionen auf einer 5-Punkte-Likert-Skala reichten von 0 („Es juckte mich nicht“) bis 4 („Es juckte mich sehr, sehr stark“) (Item 1). Zusätzlich wurde gefragt, ob der Juckreiz das Einschlafen erschwerte und ob der Juckreiz zum nächtlichen Aufwachen führte (jeweils mit „Ja“ oder „Nein“ zu beantworten; Item 2). Am Abend erfolgte eine weitere Abfrage zum Juckreiz am Tag, mit denselben Antwortoptionen der morgendlichen Erhebung. Falls die Betroffenen Juckreiz verspürten (Antwortoptionen 1–4) sollte zudem eine Angabe zum Kratzverhalten gemacht werden, mit 5 Antwortoptionen von „Kein Kratzen oder Rubbeln“, bis „Kratzen oder Rubbeln bis die Haut blutete“ (Item 3).

Beim ItchRO(Obs) wurde die jeweilige betreuende Person täglich am Morgen nach der Schwere der Juckreiz-assoziierten Symptome, wie Rubbeln, Kratzen, Hautverletzungen, Schlafstörungen oder Reizbarkeit in der Nacht, gefragt und sollte eine Einstufung von 0 („Nicht beobachtet oder berichtet“) bis 4 („Sehr schwerwiegend“) vornehmen (Item 1). Darauf folgend wurde abgefragt, was zu der Einstufung führte (z. B. Bericht des Kindes, beobachtetes Kratzen/Hautveränderungen oder Schlafstörungen) (Item 2). Zusätzlich sollte die betreuende Person eine Einschätzung dazu abgeben, wie viel Zeit das Kind mit Kratzen oder Jucken beschäftigt war, von „Nicht beobachtet“ bis „Fast die gesamte Zeit/konstant“ (Item 3). Die Abfrage am Abend erfolgte analog für alle 3 genannten Items der Erhebung am Morgen, nur, dass nach dem Zeitraum vom Aufwachen bis zum zu Bett gehen gefragt wurde.

Die beiden Versionen des Instruments werden separat ausgewertet. Die primäre Auswertung erfolgt auf Basis des wöchentlichen Durchschnittswerts des ItchRO(Obs), basierend auf Item 1. Dabei wird gemäß Protokoll der höhere Wert an jedem Tag als Tageswert berücksichtigt, der den schlimmsten Juckreiz innerhalb der vergangenen 24 Stunden abbilden soll. Gemäß SAP sind folgende weitere Auswertungen der Pruritus-Werte vorgesehen:

- Wöchentlicher Durchschnittswert für den Schweregrad am Morgen
- Wöchentlicher Durchschnittswert für den Schweregrad am Abend
- 4-Wochen-Durchschnitt für den Schweregrad am Morgen
- 4-Wochen-Durchschnittswert für den Schweregrad am Abend
- Wöchentlicher Durchschnittswert für die Häufigkeit am Morgen (nur ItchRO(Obs))
- Wöchentlicher Durchschnittswert für die Häufigkeit am Abend (nur ItchRO(Obs))
- 4-Wochen-Durchschnittswert für die Häufigkeit am Morgen (nur ItchRO(Obs))
- 4-Wochen-Durchschnittswert für Häufigkeit am Abend (nur ItchRO(Obs))
- Wöchentlicher durchschnittlicher Schweregrad (basierend auf dem täglichen Durchschnitt der Erhebungen am Morgen und am Abend) (nur ItchRO(Obs))
- Wöchentlicher durchschnittlicher Schweregrad (basierend auf dem Tagesmaximum der Morgen- und Abendwerte) (nur ItchRO(Obs))

Wöchentliche Durchschnittswerte werden für Baseline und die Wochen 3, 6, 12, 18, 22, 28, 38 und 48 gebildet und zusätzlich für die randomisierte Absetzphase (RW-Phase) zu Woche 19, 20 und 22. Ab Woche 62 erfolgt die Auswertung alle 12 Wochen.

Der wöchentliche Wert entspricht dem Mittelwert der 7 Tage vor der jeweiligen Visite, der gebildet wird falls mindestens 4 der 7 täglichen Werte vorliegen. Dementsprechend ist der 4-wöchentliche Wert der Mittelwert der 28 Tage vor der jeweiligen Visite.

Zum Screening wurden die Studienteilnehmenden inklusive betreuende Personen im Umgang mit dem elektronischen Tagebuch geschult.

Als Responder werden Personen mit einer durchschnittlichen wöchentlichen Reduktion im ItchRO um mindestens 1 Punkt im Vergleich zu Baseline gewertet.

Im Fall der Studie MRX-801 wurde der ItchRO(Obs) eingesetzt, jedoch ohne die Abfrage der Häufigkeit. Primär sollte der wöchentliche Durchschnitt des Werts am Morgen berichtet werden.

### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Auf die geplanten und durchgeführten Auswertungen wird detailliert in Kapitel 2.4 eingegangen.

### Patientenrelevanz

Pruritus zählt zu den häufigsten und belastenden Symptomen des ALGS. Er kann zu offenen Wunden sowie Hautschäden führen und wird in der Literatur zudem als Grund für chirurgische Eingriffe oder Lebertransplantationen angegeben [9,10].

In der Studie ICONIC war ein dokumentierter Pruritus im ItchRO zudem ein Einschlusskriterium. In der vorliegenden Studie ist zusätzlich eine Bewertung der Symptomatik der Studienteilnehmenden durch die betreuende Person vorgesehen. Grundsätzlich wird die Selbsteinschätzung der betroffenen Person zu ihrer wahrgenommenen Symptomatik im Vergleich zu einer Fremdeinschätzung für die Nutzenbewertung bevorzugt. Aufgrund des mehrheitlich sehr jungen Alters der Testpersonen gestaltet sich die Erhebung in Form einer Selbsteinschätzung zumindest für einen Teil der Studienpopulation schwierig, sodass eine Einschätzung durch eine andere Person im Rahmen der Nutzenbewertung als adäquat erachtet wird.

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

### Validität

Das Instrument wurde in einem iterativen Prozess entwickelt, mit Literatur-Reviews sowie dem Input von klinischen Expertinnen und Experten und der FDA [9]. Kern des Entwicklungsprozesses waren Interviews mit Kindern und Erwachsenen mit ALGS oder progressiver familiärer intrahepa-

tischer Cholestase (PFIC) im Alter von 4 bis 30 Jahren sowie mit Betreuungspersonen von Patientinnen und Patienten im Alter von 5 Monaten bis 14 Jahren. Somit wurde sichergestellt, dass das Instrument für Patientinnen und Patienten relevante Items enthält, und die standardisierte Anwendbarkeit bei Kindern ab 9 Jahren wurde überprüft. Zu weiteren Aspekten der Validität finden sich hier keine Angaben. Auch zu der in der Hauptpublikation zitierten klinischen Relevanzschwelle für Veränderungen von  $\geq 1$  Punkt für das Item 1 lassen sich keine Publikationen identifizieren.

In einer Auswertung der ITCH-Studie [12] mit 37 Patientinnen und Patienten mit ALGS im Alter von 1 bis 17 Jahren fanden sich leichte (statistisch nicht-signifikante) Korrelationen des Item 1 zur Pruritus-Schwere des ItchRO(Obs) mit der klinischen Kratzskala ( $r = -0,22$ ) sowie dem PedsQL ( $r = -0,16$ ), aber keine Korrelation mit dem sBA-Wert ( $r = -0,08$ ). Moderate Korrelationen sind für die PedsQL-Fatigue ( $r = -0,58$ ) sowie das Family Impact Module des PedsQL ( $r = -0,50$ ) beschrieben. Allerdings war hier die Spanne der möglichen ItchRO-Werte eingeschränkt, da nur Personen mit einem Wert von mindestens 2 in die Studie eingeschlossen wurden.

Aufgrund der genannten Einschränkungen der berichteten Daten aus der ITCH-Studie, werden im Folgenden zusätzlich längsschnittliche Daten aus der Studie ICONIC zum ItchRO berücksichtigt. Es ist kritisch anzumerken, dass zu diesem Aspekt der Validität keine externen Daten vorliegen.

In einer Auswertung der Studie ICONIC fanden Kamath et al. bei Personen, die zu Woche 48 als Responder auf dem ItchRO(Obs) Item 1 eingestuft wurden (Veränderungen gegenüber Baseline um  $\geq 1$  Punkt;  $N = 20$ , 75 %), positive Veränderungen in allen PedsQL-Modulen (mehr als 4 bis 5 Punkte) [11]. Im Vergleich dazu fielen die Veränderungen in der Lebensqualität bei den Non-Respondern im ItchRO zu Woche 48 ( $N = 7$ ; 25 %) deutlich geringer aus bzw. waren nicht vorhanden. Allerdings waren die Unterschiede der Veränderungen zwischen Respondern und Non-Respondern bei Kontrolle der Baseline-Werte in der Lebensqualität für den PedsQL und die PedsQL-Fatigue nicht signifikant. Die Schätzer der Regressionsanalysen hatten aufgrund der kleinen Stichprobe und hohen Variabilität weite Konfidenzintervalle. Das Responsekriterium ist hier als problematisch einzuschätzen, da nicht automatisch eine Äquivalenz der klinischen Bedeutsamkeit einer Veränderung von 1 Punkt bei unterschiedlichen Ausgangswerten angenommen werden kann. Trotz dieser Einschränkungen wird hier eine Tendenz zu besserer Lebensqualität bei positiven Veränderungen im Pruritus erkennbar.

Ferner berichtet die FDA im Integrated Review von einem Dossier zur Entwicklung des ItchRO, welches der pU für die Nutzenbewertung nicht vorgelegt hat [2].

Zusammenfassend liegen Hinweise auf die Inhalts- und Konstruktvalidität des ItchRO vor. Auf Basis der vorliegenden Daten kann die Validität des Instruments nicht abschließend bewertet werden. Da die Bewertung der Pruritus-Schwere anhand einer eindimensionalen, einfachen Skala erfolgt, werden die Ergebnisse in der Nutzenbewertung berücksichtigt. Hinsichtlich der klinischen Relevanz ist die gewählte Responseschwelle von  $\geq 1$  Punkt inhaltlich nachvollziehbar. Allerdings ist unklar, ob bei der gewählten Skalierung des Items 1 einer Reduktion höherer Schweregrade (Werte von 3 oder 4) die gleiche Bedeutung zukommt wie einer Reduktion von 1 Punkt auf 0 Punkte. Für Werte  $\leq 1$  ist davon auszugehen, dass fast keine Belastung durch den Pruritus vorliegt.

Keine Validierungsdaten liegen für die weiteren Items des ItchRO vor, wie z. B. der Frequenz des Juckens (Item 3 des ItchRO(Obs)). Da die Frequenz des Juckens zudem als Teilaspekt der Pruritus-Schwere (Item 1 – Rubbeln/Kratzen) aufgefasst wird, werden diese Ergebnisse nicht gesondert dargestellt. Auch für den 4-Wochen-Durchschnittswert existieren keine Validierungsdaten. Da dieser Zeitraum zugleich der gesamten Dauer der RW-Phase entspricht, werden diese Auswertungen aufgrund potentieller Carry-over-Effekte ebenfalls nicht dargestellt.

### **Eindruck der Veränderung (PIC/CIC)**

Der Endpunkt „Eindruck der Veränderung (PIC/CIC)“ wird in der Nutzenbewertung nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Die Symptome „Juckreiz bzw. Kratzen“ wurden zusätzlich zum Tagebuch ItchRO mit einer Abfrage zum „Eindruck der Veränderung“ erhoben. Beim patientenberichteten Eindruck der Veränderung (Patient Impression of Change (PIC)) bewerteten die Patientinnen und Patienten auf einer 7-Punkte-Skala ihre wahrgenommene Veränderung des Juckreizes im Vergleich zum Juckreiz vor Beginn der Behandlung mit der Studienmedikation. Der PIC sollte bei allen Personen  $\geq 9$  Jahre erhoben werden.

Beim betreuerberichteten Eindruck der Veränderung (Caregiver Impression of Change (CIC)) bewertete die betreuende Person die Veränderung der Juckreiz-assoziierten Symptome im Vergleich zum Beginn der Behandlung mit der Studienmedikation auf einer 7-Punkte-Skala. Der CIC sollte bei allen Studienteilnehmenden eingesetzt werden, unabhängig vom Alter.

Die eingesetzte Skala reichte von 1 („Viel besser“) bis 7 („Viel schlechter“), wobei niedrigere Werte eine verbesserte Symptomatik bedeuten.

Neben einer kontinuierlichen Auswertung des Endpunkts über den Mittelwert, wurden Responderanalysen mit allen Personen durchgeführt, bei denen sich eine Verbesserung im Vergleich zu Baseline gezeigt hat, entsprechend einem PIC/CIC von  $\leq 3$ . Personen mit fehlenden Werten zum jeweiligen Auswertungszeitpunkt werden als Non-Responder gewertet.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen. Siehe auch Endpunkt „ItchRO“.

Fraglich erscheint dabei die gewählte Altersgrenze von 9 Jahren, ab der eine Selbsteinschätzung in der Studie ICONIC vorgesehen war. Eine valide Symptombewertung erscheint zumindest für Kinder im schulfähigen Alter möglich.

#### Validität

Informationen zur Validität der Skalen konnten weder aus den Unterlagen des pU noch mittels orientierender Literaturrecherche identifiziert werden, sodass eine abschließende Bewertung nicht getroffen werden kann.

Abhängig vom Auswertungszeitpunkt geht die Abfrage der Symptomveränderung mit einer unterschiedlich langen Erinnerungsperiode einher. In der Studie ICONIC sind dies mindestens 18 Wochen, aber auch nach z. B. 48 Wochen ist noch eine Abfrage und Auswertung vorgesehen. Eine solch lange Erinnerungsperiode birgt die Gefahr eines Erinnerungsbias (Recall Bias), weshalb die Auswertungen des Tagebuchs ItchRO zur Evaluierung des Pruritus präferiert werden [13,29,30]. Auch zur Vermeidung einer Doppelerfassung werden die Ergebnisse des PIC/CIC für die vorliegende Nutzenbewertung nicht berücksichtigt.

### **Klinische Kratzskala**

Der Endpunkt „Klinische Kratzskala“ wird in der Nutzenbewertung ergänzend dargestellt.

#### Operationalisierung

##### *Beschreibung*

Pruritus wurde in den Studien ICONIC und MRX-801 auch vom klinischen Prüfpersonal anhand von Beobachtungen bei den Visiten hinsichtlich des Kratzens und der sichtbaren Schäden der Haut aufgrund des Kratzens erfasst. Die Erhebung erfolgte über eine klinische Kratzskala mit folgenden 5 Punkten:

- 0 = Kein Pruritus
- 1 = Reiben oder leichtes Kratzen, wenn keine Ablenkung erfolgt
- 2 = Aktives Kratzen ohne offensichtliche Hautabschürfungen
- 3 = Abrasion offensichtlich
- 4 = Hautverstümmelung, Blutungen und Narbenbildung offensichtlich

Wenn möglich, sollte dieselbe Person die Bewertungen zu den Visiten einer Testperson vornehmen.

Ausgewertet wurde das Instrument in der Studie ICONIC sowohl kontinuierlich über die Veränderung von Baseline zum jeweiligen Auswertungszeitpunkt als auch über Responderanalysen. Als Responderkriterien wurden eine Verringerung um mindestens 1 Punkt und um mindestens 2 Punkte gewertet. Im Fall fehlender Werte wurde die Person als Non-Responder gewertet.

Für Studie MRX-801 konnte kein Auswertungsplan identifiziert werden.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen. Siehe auch Endpunkt „ItchRO“.

#### Validität

Die klinische Kratzskala wurde ursprünglich zur Beurteilung von Juckreiz vor und nach einem chirurgischen Eingriff bei Kindern mit ALGS und PFIC eingesetzt [35]. Der Publikation konnten jedoch keine Informationen zur Entwicklung oder der formalen Validierung entnommen werden. Kritisch ist zu sehen, dass das Instrument bei den Antwortoptionen das Verhalten (Kratzen) und Oberflächencharakteristika der Haut (Wunden/Narben) vermischt. Auch ist fraglich, wie mit bereits (zu Baseline) bestehenden Hautverstümmelungen und Narben umgegangen wurde; ob diese dokumentiert wurden oder möglicherweise in Folgebewertungen eingehen. Die Validität der klinischen Kratzskala ist somit unklar. Auch konnte keine Begründung für die gewählten Responsekriterien identifiziert werden.

Grundsätzlich wird die Selbsteinschätzung der betroffenen Person zu ihrem Krankheitszustand im Vergleich zu einer Fremdeinschätzung für die Nutzenbewertung bevorzugt. Aufgrund des mehrheitlich sehr jungen Alters der Testpersonen gestaltet sich die Erhebung in Form einer Selbsteinschätzung zumindest für einen Teil der Studienpopulation schwierig, sodass eine Einschätzung durch eine dritte Person im Rahmen der Nutzenbewertung als adäquat erachtet wird. Hierbei wird prinzipiell eine Einschätzung durch die betreuende Person bevorzugt, da von einem engen Patienten-Betreuer-Verhältnis auszugehen ist.

Mit der Fremdbewertung durch das Prüfpersonal liegt jedoch eine weitere Bewertung des Schweregrads aus klinischer Perspektive vor, in die außerdem im Vergleich zur Betreuerbewertung verstärkt Hauerletzungen mit in die Bewertung eingehen. Dennoch besteht mit der

klinischen Kratzskala eine gewisse Doppelerfassung des Symptoms „Pruritus“ und der Endpunkt wird lediglich ergänzend berücksichtigt.

Die Validität der Einschätzungen bei Säuglingen, der Patientenpopulation der Studie MRX-801, ist darüber hinaus unklar, da die Fähigkeit sich zu Kratzen bei Säuglingen erst motorisch ausgebildet werden muss und hier mit großen Entwicklungsschwankungen in den ersten Lebensmonaten zu rechnen ist.

### **2.3.2.2 sBA**

Der Endpunkt „sBA“ wird in der Nutzenbewertung aufgrund fehlender Patientenrelevanz ergänzend dargestellt, da es sich um den primären Endpunkt der Studie ICONIC handelt.

#### Operationalisierung

##### *Beschreibung*

Die mittlere Veränderung des nüchternen Serumgallensäure(sBA)-Spiegels zu Woche 22 gegenüber Woche 18 bei Personen, die zuvor auf die Maralixibat-Behandlung angesprochen hatten, definiert als sBA-Reduktion  $\geq 50\%$  zu Woche 12 oder Woche 18 gegenüber Baseline, war der primäre Endpunkt der Studie ICONIC. Darüber hinaus waren die Veränderung des sBA-Spiegels von Baseline zu den jeweiligen Erhebungszeitpunkten sowie von Woche 18 zu Woche 22 in der gesamten Studienpopulation als weitere Endpunkte spezifiziert.

Der sBA-Spiegel wurde im Rahmen der Lebewerte als Laborparameter über eine Blutprobe erhoben. An den entsprechenden Studienvisiten mussten die Patientinnen und Patienten vor der Blutentnahme mindestens 4 Stunden lang fasten (nur Wasser war erlaubt).

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Beim sBA-Spiegel handelt es sich um einen Laborparameter, der, basierend auf einer Blutprobe, bei jeder Studienvisite gemessen wird. Der pU beschreibt sBA als einen Cholestase-Biomarker und stellt die Ergebnisse in Modul 4 ergänzend beim Endpunkt „Pruritus“ dar.

Dieser ergänzenden Darstellung beim Endpunkt „Pruritus“ wird nicht gefolgt. Es wurden keine Daten eingereicht, die den sBA-Spiegel als Surrogat für Pruritus qualifizieren. Eine Publikation auf Basis der ITCH-Studie, einer weiteren Studie mit Maralixibat in einer nicht fachinformationskonformen Dosierung, konnte kein Zusammenhang zwischen dem betreuerberichteten ItchRO bzw. der klinischen Kratzskala zur Erfassung des Pruritus und dem sBA-Spiegel zeigen [12].

Die Eignung des sBA-Spiegels als Parameter zur Therapiesteuerung erscheint ebenfalls fraglich. Zum einen da insbesondere schwerer Pruritus und Leberversagen die Patientinnen und Patienten mit ALGS dazu veranlasst, sich einer chirurgischen Intervention oder Lebertransplantation zu unterziehen [14]; eine Orientierung am sBA-Spiegel ist dabei nicht beschrieben. Zum anderen ist der sBA-Wert bei den Personen aus dem GALA-Register zu einem Großteil als fehlend angegeben, mit der Begründung, dass eine häufige Messung des sBA-Werts in der klinischen Praxis nicht erfolgt [23]. Da es sich bei der Veränderung des sBA-Spiegels von Woche 18 bis Woche 22 in der mITT-Population jedoch um den primären Endpunkt der Studie ICONIC handelt, werden die Ergebnisse ergänzen im Anhang dargestellt.

#### Validität

Aufgrund der unklaren Patientenrelevanz wird die Validität des Endpunkts nicht beurteilt.

### **2.3.2.3 Körperliche Entwicklung**

Der Endpunkt „Körperliche Entwicklung“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Körpergröße und Körpergewicht als Parameter der körperlichen Entwicklung wurden im Rahmen einer physischen Untersuchung bei jeder Visite erhoben und nach Alter und Geschlecht standardisiert. Für Kleinkinder unter 24 Monaten wurden die Wachstumskurven der WHO [36] zur Ableitung der z-Scores herangezogen und ab einem Alter von 24 Monaten die Wachstumskurven des Centers for Disease Control and Prevention (CDC) [3].

In der Studie MRX-801 ist die Messung weiterhin dahingehend spezifiziert, dass sie nach standardisierten Vorgaben erfolgen soll, die den Studienzentren zur Verfügung gestellt werden. Die Größe (Länge) wird mit einem geeichten Kopfteil durch zwei unabhängige Messungen auf 0,1 cm genau gemessen (und durch eine dritte Messung, wenn die Werte um mehr als 0,5 cm abweichen). Das Gewicht wird mit einer geeichten Säuglingswaage durch zwei unabhängige Messungen ermittelt und auf 0,1 kg genau aufgezeichnet (und durch eine dritte Messung, wenn die Werte um mehr als 0,3 kg abweichen). Für die Studie ICONIC sind entsprechende Angaben nicht dokumentiert.

##### *Bewertung*

Insgesamt ist unklar, ob die Bestimmung der Parameter in der Studie ICONIC standardisiert (z. B. Stehhöhe, Mittelwert aus mehreren Messungen) von geschultem Personal und jeweils zur selben Tageszeit erfolgte. Die Patientinnen und Patienten mussten jedoch 4 Stunden vor jeder Visite nüchtern bleiben. Für Studie MRX-801 ist die Operationalisierung nachvollziehbar.

#### Patientenrelevanz

Wachstumsstörungen treten häufig bei Kindern mit ALGS auf [10,14]. Körpergewicht und Körpergröße, bzw. Abweichungen davon, werden daher im vorliegenden Anwendungsgebiet als relevante Parameter eingestuft. Dabei werden Angaben, welche für Alter und Geschlecht adjustiert wurden, gegenüber absoluten Werten bevorzugt herangezogen.

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

#### Validität

Wie bereits bei der Operationalisierung angemerkt, ist in den Studienunterlagen der Studie ICONIC nicht beschrieben, ob die Ermittlung von Körpergröße und Körpergewicht standardisiert erfolgte. Eine Schwankungsanfälligkeit der Messungen ist nicht ausgeschlossen und die Validität der Erhebung kann nicht abschließend bewertet werden.

Die Auswertung erfolgt über Mittelwertunterschiede im Vergleich zu Baseline. Vom pU wurde keine Evidenz zu einer Relevanzschwelle hinsichtlich der Veränderungen vorgelegt.

### 2.3.2.4 Xanthomatose

#### **Xanthom-Skala**

Der Endpunkt „Xanthom-Skala“ wird in der Nutzenbewertung aufgrund nicht-gegebener Validität und unklarer Patientenrelevanz nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Xanthome wurden durch das Prüfpersonal oder einen geeigneten Beauftragten mittels einer Xanthom-Skala bewertet. Die Werte der Skala können von 0 bis 4 Punkte reichen. Dabei steht 0 für „Keine Anzeichen von Xanthomatose“, 1 für „Weniger als 20 verstreute einzelne Läsionen“, 2 für „Mehr als 20 Läsionen, die die Aktivitäten nicht beeinträchtigen oder einschränken“, 3 für eine „Große Anzahl von Läsionen, die aufgrund ihrer großen Anzahl oder Größe eine Entstellung des Gesichts oder der Extremitäten verursachen“, und 4 für „Xanthome, die aufgrund ihrer übermäßigen Größe oder Anzahl die Funktion beeinträchtigen (z. B. den Gebrauch der Hände oder die Fähigkeit zu gehen)“. Die Erhebung erfolgt zu Baseline, zu Woche 18 und Woche 48 sowie im Verlauf der optionalen Behandlungsverlängerung.

##### *Bewertung*

Die Operationalisierung ist weitgehend nachvollziehbar. Es konnten keine Informationen dazu identifiziert werden, ob eine Schulung des klinischen Personals zur Identifikation und Bewertung von Läsionen vorgesehen war.

#### Patientenrelevanz

Xanthome sind flecken- oder knotenartige Ablagerungen von Cholesterin unter der Haut. Sie treten bei ungefähr 30 bis 40 % aller Patientinnen und Patienten mit ALGS auf und das mehrheitlich im frühen Kindesalter [10]. Abhängig von der Ausprägung und Schwere der Betroffenheit, können Xanthome für die Patientinnen und Patienten sehr belastend sein, gerade wenn sie an Händen, Füßen oder im Gesicht auftreten. In einigen Fällen sind sie unter anderem als Grund für die Durchführung von chirurgischen Eingriffen oder Lebertransplantation angegeben [15]. Xanthome, die die betroffenen Personen entstellen sind dabei vorrangig genannt. In der vorliegenden Skalierung der Xanthom-Skala entspricht dies den Ausprägungen 3 und 4. Unklar ist jedoch, inwiefern einzeln auftretende Läsionen oder solche ohne Beeinträchtigung für die Patientinnen und Patienten und insbesondere Kleinkinder relevant sind. In der vorliegenden Operationalisierung wird die Erfassung der Xanthome deshalb als unklar eingestuft.

#### Validität

Die in der Studie ICONIC eingesetzte Xanthom-Skala wurde erstmals in einer Studie zur partiellen externen Gallengangsdiversion bei Personen mit ALGS eingesetzt [4]. Informationen zu Entwicklung oder Validierung des Instruments sind jedoch weder dieser Publikation noch den Studienunterlagen zu entnehmen. Weiterhin geht aus den Studienunterlagen nicht hervor, dass die Erhebung standardisiert erfolgte bzw. das Studienpersonal hinsichtlich der Erhebung geschult wurde. Insbesondere die Abgrenzung zwischen den Werten 2 und 3 der Skala erscheint einer gewissen Subjektivität zu unterliegen. In Verbindung mit fehlenden Daten zu Inter- und Intrarater-Reliabilität erscheint fraglich, inwiefern im vorliegenden Setting eine valide Erhebung des Endpunkts möglich ist.

### **Veränderung der Xanthome**

Der Endpunkt „Veränderung der Xanthome“ wird in der Nutzenbewertung aufgrund unklarer Patientenrelevanz und unklarer Validität nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Die Veränderung des Schweregrads von Xanthomen im Vergleich zu Baseline wird in der Studie ICONIC von den betreuenden Personen auf einer Skala von 1 (= viel besser) bis 7 (= viel schlechter) eingeschätzt. Erhebungen waren zu Woche 18, 22, 48, 84 und 98 geplant.

Die Auswertung erfolgt über Responderanalysen, wobei alle Personen mit einem Wert  $\leq 3$  (entspricht viel besser, besser und etwas besser) als Responder gewertet werden. Analysen ausschließlich für die RW-Phase waren nicht vorgesehen.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Xanthome können abhängig von der Operationalisierung und dem Ausmaß der Betroffenheit patientenrelevant sein (siehe Endpunkt „Xanthom-Skala“). Mit der vorliegenden Veränderungsabfrage bei den betreuenden Personen wird jedoch nicht die Belastung der Patientinnen und Patienten durch Xanthome abgefragt, sondern der Schweregrad der Xanthome. Die Erfassung einer patientenrelevanten Verbesserung setzt insofern eine relevante Beeinträchtigung der Patientinnen und Patienten zu Behandlungsbeginn voraus. Die Belastung und Beeinträchtigung, die mit einzelnen Läsionen einhergeht, ist unklar und demnach ist auch die Patientenrelevanz des Endpunkts in der vorliegenden Operationalisierung unklar.

#### Validität

Es konnten keine Informationen zur Validität der eingesetzten Skala bzw. des berücksichtigten Responderkriteriums von  $\leq 3$  Punkten identifiziert werden. Auch ist nicht begründet, weshalb ausschließlich Verbesserungen und keine Verschlechterungen berücksichtigt wurden. Die vorliegende Veränderungserfassung kommt ohne eine Erhebung zu Baseline aus und es ist auch keine anderweitige Dokumentation der Xanthom-Betroffenheit zu Behandlungsbeginn vorgesehen, z. B. durch eine Fotodokumentation. Da die weiteren Erhebungen erst in deutlichem Abstand (mindestens 18 Wochen) festgesetzt sind, ist ein deutlicher Erinnerungsbias (Recall Bias) nicht ausgeschlossen. Insgesamt wird die Validität des Endpunkts als unklar bewertet.

### **2.3.2.5 Leberfunktion**

Der Endpunkt „Leberfunktion“ mittels Alanin-Aminotransferase (ALT), alkalischer Phosphatase (ALP), Gesamtbilirubin und direktem Bilirubin wird in der Nutzenbewertung aufgrund nicht-gegebener Patientenrelevanz nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

In den Studien ICONIC und MRX-801 wurden die Leberwerte ALT, ALP, Gesamtbilirubin und direktes Bilirubin als Endpunkt für die Leberfunktion ausgewertet.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

Biochemische Leberparameter sind kein unmittelbar patientenrelevanter Endpunkt. Die Validität der Leberparameter als Surrogat patientenrelevanter Morbiditätsendpunkte wird weder vom pU hergeleitet noch durch entsprechende Literatur ausreichend gestützt. Die Patientenrelevanz wird aus diesem Grund als nicht gegeben gewertet.

### Validität

Aufgrund der fehlenden Patientenrelevanz wird die Validität des Endpunkts nicht beurteilt.

#### **2.3.2.6 PedsQL-Fatigue**

Der Endpunkt „PedsQL-Fatigue“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Zusätzlich zum PedsQL wurde in der Studie ICONIC die PedsQL – Multidimensional Fatigue Scale (PedsQL-Fatigue), Version 3.0, eingesetzt; einem Zusatzmodul zur Messung der Fatigue und Fatigue-assoziiierter Belastungen bei pädiatrischen Patientinnen und Patienten.

Auch die PedsQL-Fatigue kam in unterschiedlichen Versionen für die Altersklassen zum Einsatz, analog zum PedsQL. Allerdings existieren hier nur Versionen ab einem Alter von 2 Jahren, weshalb dieses Instrument nur bei einer Subgruppe der Patientinnen und Patienten zum Einsatz kam. Die PedsQL-Fatigue besteht aus 18 Items, verteilt auf die folgenden 3 Domänen mit je 6 Items:

- Generelle Fatigue
- Schlaf-/Ruheprobleme
- Kognitive Fatigue

Analog zum PedsQL erfolgte die Beantwortung der Items auf einer 4-Punkte-Likert-Skala von 1 (= nie; entspricht der besten Funktion) bis 4 (= fast immer; entspricht der schlechtesten Funktion). Davon ausgenommen ist der Selbstbericht der jungen Kinder (5–7 Jahre), in dem die Beantwortung auf einer 3-Punkte-Likert-Skala von 0 (= überhaupt nicht) über 2 (= manchmal) bis 4 (= viel) vorgesehen ist. Die Werte werden anschließend in eine Skala von 1 bis 100 invertiert. Skalenwerte werden nur berechnet, sofern mindestens 50 % der zugehörigen Items beantwortet wurden. Höhere Werte deuten auf eine geringere Belastung durch die Fatigue hin. Die Bezugsperiode der Items beträgt in den vorliegenden Versionen ein Monat. Basierend auf dem Alter zur Baseline wurde dieselbe Version während der gesamten Dauer der Studie verwendet. Die Auswertungen erfolgten getrennt für die Selbsteinschätzungen der Kinder und die Fremdeinschätzungen durch die betreuenden Personen.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

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

### Validität

Die Validität der PedsQL-Fatigue wurde in einem vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im zugrundeliegenden Anwendungsgebiet als valide eingeschätzt.

### 2.3.3 Lebensqualität

#### **PedsQL**

Der Endpunkt „PedsQL“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Zur Erfassung der gesundheitsbezogenen Lebensqualität in der Studie ICONIC wurde der generische Pediatric Quality of Life Inventory (PedsQL) in der Version 4.0 eingesetzt. Dabei handelt es sich um ein multidimensionales Instrument zur Selbst- und Fremdeinschätzung für Kinder und Eltern zur Beurteilung der Lebensqualität bei Kindern, Jugendlichen und jungen Erwachsenen im Alter von 1 Monat bis 25 Jahren [17]. In Abhängigkeit vom Alter der Testperson wurden in der Studie ICONIC unterschiedliche Versionen des PedsQL eingesetzt: Patienten- und Elternberichte in je einer Version für 5- bis 7-Jährige, 8- bis 12-Jährige und 13- bis 18-Jährige sowie Elternberichte für Kleinkinder in je einer Version für 1–12 Monate, 13–24 Monate und 2–4 Jahre.

Je nach Version umfasst das Instrument 23 Items (alle Versionen ab 2 Jahre), 36 Items (1–12 Monate) oder 45 Items (13–24 Monate), aufgeteilt auf die folgenden Domänen:

- Physische Funktion
- Physische Symptome (nur in den Versionen bis 24 Monate)
- Emotionale Funktion
- Soziale Funktion
- Schulische Funktion

Im Elternbericht für Kleinkinder besteht die Domäne „Schulische Funktion“ aus lediglich 3 Items, die nur beantwortet werden sollten, falls das Kind zur Schule geht oder in einer Kindertagesstätte ist.

Die Beantwortung der Items erfolgt auf einer 4-Punkte-Likert-Skala von 1 (= nie; entspricht der besten Funktion) bis 4 (= fast immer; entspricht der schlechtesten Funktion). Eine Ausnahme ist der PedsQL für den Selbstbericht der jungen Kinder (5–7 Jahre), in dem die Beantwortung auf einer 3-Punkte-Likert-Skala von 0 (= überhaupt nicht) über 2 (= manchmal) bis 4 (= viel) erfolgt. Die Werte werden anschließend in eine Skala von 1 bis 100 invertiert. Skalenwerte werden nur berechnet, sofern mindestens 50 % der zugehörigen Items beantwortet wurden. Höhere Werte deuten auf eine höhere Lebensqualität hin. Die Bezugsperiode der Items beträgt in den vorliegenden Versionen ein Monat. Basierend auf dem Alter zu Baseline wurde dieselbe Version während der gesamten Dauer der Studie verwendet. Die Auswertungen erfolgten getrennt für die Selbsteinschätzungen der Kinder und die Fremdeinschätzungen durch die betreuenden Personen.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

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

#### Validität

Die Validität des PedsQL wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet, allerdings nur in den Versionen ab 2 Jahren.

Seit 2011 existieren auch Eltern-/Proxy-Versionen des PedsQL für Kleinkinder im Alter von 1 Monat bis 24 Monaten. Diese wurden in einer Studie mit 683 Eltern von gesunden, akut kranken und chronisch kranken Kindern evaluiert und zeigten eine hohe interne Konsistenz, Known-Groups-Validität und eine mit der Konzeption übereinstimmende Faktorstruktur [34].

Eine gute Validität dieser Kleinkind-Skalen zeigte sich auch in einer Studie mit über 1.000 Kindern mit wiederkehrender Mittelohrentzündung (Otitis media), deren Lebensqualität mit gesunden Kindern verglichen wurde [8]. Bei Kindern mit einer chronischen Erkrankung der Leber zeigte sich im Alter 1–12 Monate im Vergleich zu gesunden Kindern eine niedrigere elternberichtete Lebensqualität, allerdings nicht in der Altersgruppe der Kinder im Alter 13–24 Monate [28].

Die ähnliche Konzeption der Skalen im Vergleich zu den Versionen für die älteren Kinder und die gute Known-Groups-Validität lassen vermuten, dass auch diese Versionen änderungssensitiv sind. Untersuchungen dazu legt der pU allerdings nicht vor und eine kurze, orientierende Recherche ergab diesbezüglich keine weiteren Ergebnisse, sodass bezüglich der Änderungssensitivität noch Unklarheit besteht.

Zusammenfassend werden der PedsQL als geeignet zur Erfassung der gesundheitsbezogenen Lebensqualität in den vorliegenden Altersgruppen bewertet.

## 2.3.4 Sicherheit

### Unerwünschte Ereignisse

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

#### Operationalisierung

##### *Beschreibung*

Unerwünschte Ereignisse (UE) sind in der Studie ICONIC definiert als jegliche ungünstige und ungewollte Anzeichen, Symptome oder Erkrankungen in zeitlichem Zusammenhang mit der Einnahme der Studienmedikation; unabhängig davon, ob das UE im Zusammenhang mit der Prüfmedikation steht. Alle UE, die von den Testpersonen, betreuenden Personen oder dem Prüfpersonal berichtet werden, sowie Labor-Abnormalitäten sollten im Case Report Form dokumentiert werden. Klinisch bedeutsame abnormale Labortestergebnisse können nach Ansicht des Prüfpersonal ein UE darstellen oder mit einem UE in Verbindung gebracht werden. Die Kodierung erfolgte anhand des Medical Dictionary for Regulatory Activities (MedDRA), Version 22.1, unter Verwendung der Systemorganklassen (SOC) und Preferred Terms (PT). Folgendes stellt kein UE dar:

- Vor Beginn der Behandlung bestehende Erkrankungen/Symptome, die nicht unerwartet fortschreiten oder bei denen sich der Schweregrad nach Verabreichung des Arzneimittels nicht ändert.
- Die untersuchte Erkrankung und/oder Anzeichen und Symptome, die mit der vorliegenden Erkrankung in Verbindung gebracht werden, wie Gelbsucht oder Juckreiz, oder Anomalien der Leberenzyme, die bereits während des Screening-Zeitraums oder bei der Erstuntersuchung vorhanden waren.
- Therapieversagen oder Mangel an Wirksamkeit.

Symptome der untersuchten Krankheit sollten nicht als UE eingestuft werden, solange sie im Rahmen der normalen täglichen Fluktuation oder des erwarteten Krankheitsverlaufs liegen und Teil der zu erhebenden Wirksamkeitsdaten sind; eine signifikante Verschlimmerung der Symptome sollte jedoch als UE erfasst werden.

Der Schweregrad von UE wurde mittels Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0, bewertet werden. Falls für das UE keine Graduierung vorliegt, sollte der Schweregrad basierend auf der folgenden Einstufung vorgenommen werden:

- Mild (leicht zu tolerieren, beeinträchtigt die täglichen Aktivitäten nicht).
- Moderat (Unwohlsein, stört die normalen täglichen Aktivitäten).
- Schwer (handlungsunfähig; führt zu erheblichen Beeinträchtigungen bei gewöhnlichen täglichen Aktivitäten).

- Im Falle eines lebensbedrohlichen, nicht graduierten UE wird dennoch eine Einstufung in CTCAE-Grad 4 vorgenommen, gleiches gilt bei UE mit Todesfolge (CTCAE-Grad 5).

Als schwerwiegendes unerwünschtes Ereignis (SUE) wurden UE definiert, die

- zum Tode führten,
- direkt lebensgefährdend waren,
- zur Hospitalisierung bzw. Verlängerung einer Hospitalisierung führten (mindestens 24 Stunden, keine elektiven Aufnahmen),
- zu einer anhaltenden oder erheblichen Behinderung oder Beeinträchtigung führten,
- zu einer kongenitalen Anomalie / angeborenen Fehlbildung führten oder
- als medizinisch bedeutsam angesehen wurden, jedoch nicht zum Tod führten, nicht lebensbedrohlich waren oder keinen Krankenhausaufenthalt erforderten, aber nach ärztlicher Beurteilung die Patientin / den Patienten gefährden und einen medizinischen oder chirurgischen Eingriff erforderten, um eines der in dieser Definition aufgeführten Ergebnisse zu verhindern. Beispiele können sein: Intensive Behandlung in der Notaufnahme oder zu Hause aufgrund eines allergischen Bronchospasmus; Blutdyskrasie oder Anfälle, die nicht zu einer Hospitalisierung führten; Entwicklung einer Medikamentenabhängigkeit und Medikamentenmissbrauch.

Gemäß SAP waren folgende UE von besonderem Interesse definiert:

- Diarrhö (PT „Diarrhö“ und PT „Gastroenteritis“)
- Ereignisse im Zusammenhang mit erhöhtem Bilirubin (PT „Bilirubin erhöht“)
- Erhöhte Transaminasen (PT „Alanin-Aminotransferase erhöht“ und PT „Aspartat-Aminotransferase erhöht“)
- Mangel an fettlöslichen Vitaminen (definierte Liste an PT, u. a. zu Vitaminmangel, Augenerkrankungen und Knochenproblemen)

Als behandlungsbezogene UE (TEAE) sind alle UE definiert, die nach Einnahme der ersten Dosis der Studienmedikation auftraten und ihren Beginn spätestens 14 Tage nach der letzten Dosis hatten. Im Fall einer Behandlungsunterbrechung werden UE, die nach 14 Tagen auftreten und vor der Wiederaufnahme der Therapie enden, nicht als behandlungsbezogen gewertet. UE werden für die gesamte Studie und zudem getrennt für die Phasen der Dosisescalation / stabilen Dosis (Wochen 0–18), der RW-Phase und der langfristigen Behandlungsphase der Studie zusammengefasst. Hierbei wurden nur in der jeweiligen Studienphase neu auftretende oder sich verschlimmernde UE gewertet.

In der Studie MRX-801 werden alle UE ab der ersten Dosis und alle SUE ab dem Zeitpunkt der Einwilligungserklärung bis zum Studienende berichtet. Dabei werden alle UE als behandlungsbezogene UE gewertet. Nicht als UE werden abnormalen Laborergebnisse oder Anzeichen und Symptome gewertet, die mit der Grunderkrankung im Zusammenhang stehen.

Die Klassifikation von schweren UE und SUE erfolgte weitestgehend analog zur Studie ICONIC. Abweichend waren folgende UE von besonderem Interesse definiert:

- Mangel an fettlöslichen Vitaminen, die eine Unterbrechung der Studienmedikation erfordern.
- Störungen der Leberparameter, die eine Unterbrechung der Studie und/oder eine Änderung der Dosis erfordern.
- Alle Ereignisse, bei denen der Verdacht besteht und/oder bestätigt wird, dass sie auf eine Propylenglykol-Toxizität zurückzuführen sind, einschließlich neurologischer Komplikationen, Hämolyse und Herzrhythmusstörungen.

### *Bewertung*

Die Operationalisierung ist weitgehend nachvollziehbar. Bei der Auswertung der UE in den unterschiedlichen Studienzeiträumen (vor Randomisierung, RW-Phase und Phase nach Randomisierung) werden die Ereignisse nur in dem Zeitraum gewertet, in dem das Ereignis begann. Bestehende UE werden nicht gewertet. Offensichtlich ist das auch bei der Auswertung nach unterschiedlichen Dosierungen der Fall, sodass UE von Personen in der Auftitrationsphase mit  $< 400 \mu\text{g}/\text{kg}/\text{Tag}$  bei der Auswertung der Dosierung  $400 \mu\text{g}/\text{kg}/\text{Tag}$  nicht enthalten sind.

Die Definition der UE von besonderem Interesse wurde nach Entblindung der Studie geändert. Ursprünglich waren gastrointestinale Ereignisse, Erkrankungen, die mit einer Verschlechterung der Leberfunktion einhergehen, fettlösliche Vitaminmangelerscheinungen und Retardierung definiert. Für dieses Vorgehen konnte keine Begründung identifiziert werden.

### Patientenrelevanz

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

### Validität

Gemäß Operationalisierung sollen Symptome, die im Zusammenhang mit der Erkrankung von Maralixibat stehen, nicht als UE erfasst werden, sofern es sich um keine signifikante Verschlimmerung handelt. Die zugrundeliegenden Symptome sind nicht definiert. Es bleibt unklar, ob mit dieser Einschränkung ausreichend zwischen Grunderkrankung und Nebenwirkungen unterschieden werden kann. Weiterhin wurde erst mit Protokollamendment 3 vom 13.11.2015, und somit nach Einschluss der letzten Person in die Studie ICONIC, im Abschnitt zur Dokumentation der UE ein Abschnitt eingefügt, nachdem Symptome der Erkrankung nicht weiter als UE klassifiziert werden sollten, sofern sie im Rahmen normaler täglichen Schwankung oder Bestandteil der erwarteten Progression sind, die bereits bei der Wirksamkeit abgebildet ist. Zuvor war bei der Definition eines UE lediglich beschrieben, dass Ereignisse der untersuchten Erkrankung und/oder Anzeichen und Symptome, die mit der vorliegenden Erkrankung in Verbindung gebracht werden, wie Gelbsucht oder Juckreiz, oder Anomalien der Leberenzyme, die bereits während des Screening-Zeitraums oder bei der Erstuntersuchung vorhanden waren, nicht als UE gewertet werden sollten. Insgesamt ist unklar, inwiefern Symptomatik auch als UE erfasst wurde.

In Modul 4 präsentiert der pU Auswertungen aller UE in den unterschiedlichen Studienphasen (z. B. RW-Phase), wobei UE über einer Dosis von  $400 \mu\text{g}/\text{kg}/\text{Tag}$  hier nicht berücksichtigt werden. Dabei bezieht sich der Anteil der Personen mit Ereignis auf diejenigen Patientinnen und Patienten, die sich jeweils noch in der Studie befanden. Dieser Darstellung wird nicht gefolgt. Für die Nutzenbewertung sind prinzipiell alle UE relevant, die bis einschließlich einer Dosis von  $400 \mu\text{g}/\text{kg}/\text{Tag}$  auftraten, unabhängig von der Studienphase. Laut SAP werden zudem nur die jeweils neu auftretenden oder sich verschlimmernden UE in einer Studienphase erfasst.

Aus diesem Vorgehen folgt, dass die UE-Ergebnisse der RW-Phase nicht aussagekräftig für die Bewertung der Sicherheit von Maralixibat im Vergleich zu Placebo sind. Die Auswirkungen des Absetzens eines Wirkstoffs nach einer längeren Behandlungsphase ist nicht mit einer Nicht-Gabe des Wirkstoffs zu vergleichen. UE, die sich bereits in den 18 vorangegangenen Behandlungswochen ereigneten, sind nicht enthalten. Insofern ist mit einer Verzerrung zugunsten des Maralixibat-Arms zu rechnen, da eine mögliche Verbesserung/Verringerung oder ein Ausbleiben der Nebenwirkungen im Placebo-Arm nicht erfasst wird.

### 2.3.5 Übersicht der Erhebungszeitpunkte

Eine Übersicht der Erhebungszeitpunkte der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte findet sich in den Tabellen 8 und 9. Die Sicherheitsendpunkte wurden bis 30 Tage nach Ende der Behandlungszeit erhoben.

*Tabelle 8: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie ICONIC*

| Studienvisite<br>(Woche)   | Screening  |           | Base-<br>line | Offene<br>Behandlungsphase <sup>1)</sup> |    |     |     | RW-<br>Phase <sup>2)</sup> | Langfristige<br>Behandlungs-<br>phase <sup>3)</sup> |     |     | Optio-<br>nale<br>LTE <sup>4)</sup> | FuP |
|----------------------------|------------|-----------|---------------|--|----|-----|-----|----------------------------|---|-----|-----|-------------------------------------|-----|
|                            | Tag<br>-28 | Tag<br>-1 |               | W3                                       | W6 | W12 | W18 |                            | W22   | W28 | W38 |                                     |     |
| Todesfälle <sup>5)</sup>   |            |           |               |  |    |     |     | X                          |   |     |     |                                     |     |
| ItchRO <sup>6)</sup>       |            |           |               |  |    |     |     | X                          |   |     |     | X                                   |     |
| Klinische Kratzskala       | X          |           | X             | X  | X  | X   | X   | X                          | X   | X   | X   | X                                   |     |
| Körperliche<br>Entwicklung | X          |           | X             | X  | X  | X   | X   | X                          | X   | X   | X   | X                                   |     |
| sBA                        | X          |           | X             | X  | X  | X   | X   | X                          | X   | X   | X   | X                                   |     |
| PedsQL-Fatigue             |            |           | X             |  |    |     | X   | X                          |   |     | X   | X                                   |     |
| PedsQL                     |            |           | X             |  |    |     | X   | X                          |   |     | X   | X                                   |     |
| Unerwünschte<br>Ereignisse |            |           |               |  |    |     |     | X                          |   |     |     |                                     |     |

<sup>1)</sup> Die Aufitrationsphase endete mit Woche 6. Am Ende der Wochen 1, 2, 4, 5 und 9 wurden zusätzlich zu den Visiten wöchentliche Telefonanrufe durchgeführt, zur Erhebung von Begleitmedikation und UE.

<sup>2)</sup> Die RW-Phase dauerte von Woche 19 bis 22. Als Baseline dieser RW-Phase zählen die Erhebungen zu Woche 18. Ein -Telefonanruf wurde zu Woche 20 durchgeführt, zur Erhebung von Begleitmedikation und UE.

<sup>3)</sup> Telefonanruf zu Woche 23–27, 33 und 43, zur Erhebung von Begleitmedikation und UE.

<sup>4)</sup> Die optionale LTE umfasst die optionale 52-wöchige Behandlungsverlängerung nach Woche 48 und die daran anschließende optionale langfristige Behandlungsverlängerung nach Woche 100. Alle Endpunkte wurden während dieser Phase in regelmäßigen Abständen erhoben. Da Ergebnisse dieser Phase nicht mehr berücksichtigt werden (siehe Kapitel 2.4), wird auf eine genaue Darstellung der Erhebungszeitpunkte verzichtet.

<sup>5)</sup> Im Rahmen der Sicherheit erfasst.

<sup>6)</sup> Das Tagebuch sollte bis Woche 48 kontinuierlich zweimal täglich ausgefüllt werden.

Abkürzungen: FuP: Follow-up; ItchRO: Itch Reported Outcome; LTE: (Langfristige) Behandlungsverlängerung; PedsQL: Pediatric Quality of Life Inventory; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; RW: Randomisierte Absetzphase; sBA: Serurgallensäure; UE: Unerwünschtes Ereignis; W: Woche.

*Tabelle 9: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie MRX-801*

| Endpunkt \ Studiervisite (Woche) | Baseline | Auftritrationsphase <sup>1)</sup> |    | Erhaltungsphase | Ende Hauptstudie <sup>2)</sup> | LTE <sup>3)</sup> |       | ET <sup>4)</sup> | FuP  |
|----------------------------------|----------|-----------------------------------|----|-----------------|--------------------------------|-------------------|-------|------------------|------|
|                                  |          | W3                                | W6 | W10             | W13                            | + 4W              | + 16W |                  | + 1W |
| Todesfälle <sup>5)</sup>         | X        |                                   |    |                 |                                |                   |       |                  |      |
| ItchRO(Obs)                      | X        | X                                 | X  | X               | X                              | X                 |       | X                |      |
| Klinische Kratzskala             | X        |                                   | X  |                 | X                              | X                 | X     | X                |      |
| Körperliche Entwicklung          | X        | X                                 | X  | X               | X                              | X                 | X     | X                |      |
| Unerwünschte Ereignisse          | X        |                                   |    |                 |                                |                   |       |                  |      |

<sup>1)</sup> In der Auftritrationsphase bis Woche 6 waren weitere Visiten zu Woche 1, 2, 4 und 5 vorgesehen, die jedoch auch telefonisch absolviert werden konnten. Nur im Fall einer Visite im Studienzentrum wurde eine körperliche Untersuchung durchgeführt und dabei die Parameter der körperlichen Entwicklung erhoben.

<sup>2)</sup> Visite zum Ende der Erhaltungsphase (Woche 7 bis 13).

<sup>3)</sup> Langfristige Behandlungsverlängerung. Falls Alter < 12 Monate: Kontaktaufnahme alle 2 Wochen, Visite alle 4 Wochen. Falls Alter ≥ 12 Monate: Visite alle 16 Wochen.

<sup>4)</sup> Ende der Behandlung. Falls die letzte Visite ≤ 8 Wochen zurückliegt, wird ausschließlich der ItchRO(Obs) (falls jünger als 12 Monate) erhoben.

<sup>5)</sup> Im Rahmen der Sicherheit erfasst.

Abkürzungen: ET: Ende der Behandlung; FuP: Follow-up; LTE: Langfristige Behandlungsverlängerung; ItchRO(Obs): Itch Reported Outcome (Observer); W: Woche.

## 2.4 Statistische Methoden

Die statistischen Methoden basieren auf dem SAP der Studie ICONIC, der in der finalen Version 1.0 auf den 17.03.2020 datiert ist und damit nach Entblindung finalisiert wurde.

Für die Studie MRX-801 liegt kein SAP vor, weshalb die Ausführungen dieses Kapitels auf die Studie ICONIC fokussieren.

### Analysepopulationen

- Intention-to-Treat (ITT)-Population:
  - ICONIC: Alle eingeschlossenen Personen, die mindestens eine Dosis der Studienmedikation erhalten haben.
  - MRX-801: Alle eingeschlossenen Personen.
- Sicherheitspopulation: Alle eingeschlossenen Personen, die mindestens eine Dosis der Studienmedikation erhalten haben (Studien ICONIC und MRX-801).
- RW-Population: Alle randomisierten Personen, die mindestens eine Dosis der Studienmedikation in der RW-Phase erhalten haben (nur Studie ICONIC).

### Datenschnitte

Eine Interimsanalyse war geplant, wenn alle Personen die Visite zu Woche 48 abgeschlossen oder die Studie abgebrochen haben. Die Auswertungen basieren auf einem separaten Interim-SAP vom 04.02.2018. Ergebnisse für diese Interimsanalyse liegen nicht vor. In Vorbereitung auf Meetings mit den Zulassungsbehörden sind gemäß finalem SAP weitere Interimsanalysen durchgeführt worden, deren Daten ebenfalls nicht vorliegen. Die berücksichtigten Analysen beziehen sich auf den finalen Datenschnitt vom 21.08.2020.

In der Studie MRX-801 war eine Interimsanalyse geplant, wenn mindestens 6 Kinder die 13-wöchige Kernstudie durchlaufen haben. Zum vorliegenden Datenschnitt vom 04.05.2022 wurde diese bereits von 7 Kindern abgeschlossen.

### **Präspezifizierte Subgruppenanalysen**

Weder im Protokoll der Studie ICONIC noch im Interim-SAP waren Subgruppenanalysen präspezifiziert. Lediglich im finalen SAP sind für „Alter“, „Baseline-sBA“, „Baseline-Gesamtbilirubin“, „Baseline-ALT“ und „Baseline-ItchRO(Obs)“ Subgruppenanalysen für die Endpunkte „ItchRO(Obs)“ und „sBA“ post hoc nach Entblindung definiert. Interaktionstests für die Ergebnisse der RW-Phase waren nicht vorgesehen. Ausgewertet wurde ausschließlich die Veränderung von Baseline bzw. Woche 18 zu den unterschiedlichen Auswertungszeitpunkten.

Dem Protokoll der Studie MRX-801 konnten keine Subgruppenanalysen entnommen werden.

### **Präspezifizierung geplanter und durchgeführter Analysen**

#### Auswertungszeiträume

Bei den Auswertungen der Studie ICONIC werden 3 Behandlungsphasen unterschieden:

- Offene-Behandlungsphase (Tag 1 bis Woche 18),
- RW-Phase (Woche 19 bis 22) und
- Langfristige Behandlungsphase (nach der RW-Phase (Woche > 22)).

Abhängig vom Endpunkt werden Auswertungen sowohl über den gesamten Studienzeitraum als auch getrennt für die Behandlungsphasen präsentiert. Für ausgewählte Analysen wird die langfristige Behandlungsphase als 2 separate Behandlungsphasen dargestellt: Nach-Randomisierungsphase zwischen Woche > 22 und Woche 48 sowie die Nach-Randomisierungsphase ab Woche > 48. Die Auswertungen zur Wirksamkeit beziehen sich dabei jeweils auf die Erhebung zu Baseline. Für die RW-Phase liegen zusätzlich Auswertungen gegenüber Woche 18 vor.

Die Auswertungen in Studie MRX-801 erfolgten ausschließlich deskriptiv.

#### ItchRO

Die Auswertungen zum ItchRO sind im Verlauf der Studie ICONIC mehrfach angepasst worden:

- Gemäß Studienprotokoll waren zur Auswertung des ItchRO wöchentliche Durchschnittswerte des maximalen Tageswerts als auch Responderanalysen vorgesehen, die jedoch nicht weiter beschrieben sind.
- Der Interim-SAP vor Entblindung fokussierte dann primär auf eine Auswertung der wöchentlichen Durchschnittswerte der Erhebungen am Morgen (nächtlicher Pruritus). Weiterhin waren Auswertungen der wöchentlichen Durchschnittswerte für die Erhebung am Abend und jeweils für die 4-wöchigen Durchschnittswerte sowie weitere Sensitivitätsanalysen vorgesehen. Als Responder gemäß diesem Interim-SAP waren Personen mit einer Reduktion im wöchentlichen Durchschnitt der Erhebungen am Morgen um  $\geq 1$  Punkt und weiterhin  $\geq 1,25$ ,  $\geq 1,33$  und  $\geq 1,5$  Punkte definiert. Angaben zur statistischen Methodik in diesem Dokument sind jedoch nur für Responderanalysen (Chi<sup>2</sup>- oder Fisher's Exact Test) gegeben.
- Die statistische Methodik für insbesondere die kontinuierlichen Auswertungen des ItchRO mittels ANCOVA (Analysis of Covariance) inklusive der Adjustierung für Kovariaten wurde erst post hoc nach Entblindung im finalen SAP festgelegt. Dort wurden zudem folgende, weitere Auswertungen des ItchRO bestimmt:

- Sekundärer Endpunkt der Studie ICONIC: Wöchentlicher Durchschnitt des maximalen Tageswerts bei Personen, die zuvor auf eine Therapie mit Maralixibat angesprochen haben (Reduktion um  $\geq 1$  Punkt zu Woche 12 oder Woche 18).
- Wöchentliche und 4-wöchentliche Durchschnittswerte des maximalen Tageswerts und des durchschnittlichen Tageswerts für alle randomisierten Personen.
- Auswertungen des Item 3 zur Frequenz des Kratzens (nur ItchRO(Obs)).
- Sensitivitätsanalysen der kontinuierlichen Auswertungen mit Imputationen über LOCF (Last Observation Carried Forward), mittels MMRM (Mixed Model Repeated Measures), anhand bestimmter Baseline-Charakteristika als Kovariaten (z. B. der Stratifikationsvariable „sBA-Responder“ sowie „sBA-Spiegel“, „BMI“, „Alter“) und unter Berücksichtigung einer Protokollverletzung.
- Kategoriale Auswertungen für die Anzahl und den Anteil der Tageswerte für die Erhebung am Morgen mit einem Wert von  $\leq 1$  Punkt.
- Die Responderanalysen einer Reduktion im wöchentlichen Durchschnitt der Erhebungen am Morgen ( $\geq 1$ ,  $\geq 1,25$ ,  $\geq 1,33$  und  $\geq 1,5$  Punkte) sind im finalen SAP dagegen nicht mehr enthalten. Diese Auswertungen für den ItchRO(Obs) finden sich dann doch im Studienbericht der Studie ICONIC.

## **Fehlende Werte und Daten-Imputation**

### Allgemein

Sofern ein Baseline-Wert (sowohl Tag 0 als auch Woche 18) fehlend ist, sollte jeweils der letzte verfügbare Wert als Baseline berücksichtigt werden.

In der Studie ICONIC werden Sensitivitätsanalysen mittels MMRM durchgeführt. Zusätzlich erfolgen Auswertungen zu Woche 18, 22, 48 und 100 mittels LOCF, wobei jeweils die letztverfügbare Beobachtung in der jeweiligen Studienphase in die Auswertung eingeht, also z. B. für die RW-Phase die letzte Beobachtung zwischen Woche 18 und Woche 22.

Fehlende Werte bei Responderanalysen werden allgemein als Non-Responder gewertet.

### ItchRO

Sofern ein wöchentlicher Durchschnittswert nicht gebildet werden konnte, wurde bei den Auswertungen mittels LOCF das letzte vorangegangene 7-Tage-Intervall berücksichtigt, für das ein wöchentlicher Wert gebildet werden konnte.

## **Einschätzung der statistischen Auswertungen von Studie ICONIC**

### Auswertungszeiträume

Für die Nutzenbewertung sind vorrangig die vergleichenden Auswertungen der 4-wöchigen RW-Phase relevant. Darüber hinaus werden die nicht-vergleichenden Ergebnisse für alle Endpunkte in der ITT-Population zu Woche 48 und ggf. zu Woche 18 und Woche 22 jeweils im Vergleich zu Baseline dargestellt. Ursprünglich war die Studie ICONIC auf 48 Wochen angelegt. Für die studienverlängernden Protokollamendments bedurfte es einer erneuten Einwilligungserklärung für die Teilnahme, die von einem Teil der Studienteilnehmenden nicht erteilt wurde (siehe Tabelle 13). Die Population der optionalen Behandlungsverlängerung entspricht nicht mehr dem ITT-Prinzip und die Rücklaufquoten in allen Wirksamkeitsendpunkten fallen dauerhaft unter 70 %, weshalb diese Ergebnisse nicht in der Nutzenbewertung berücksichtigt werden können.

### Analysepopulation

In den Studienunterlagen des pU zur Studie ICONIC wird eine mITT-Population als primäre Analysepopulation beschrieben. Da hier ausschließlich eine Teilpopulation mit einer Reduktion von  $\geq 50\%$  im sBA-Wert zu Woche 12 oder 18 in die Auswertungen eingeht, werden diese Ergebnisse in der Nutzenbewertung nicht berücksichtigt, abgesehen vom primären Endpunkt, welcher ergänzend dargestellt wird (siehe Anhang). Ferner ist zu berücksichtigen, dass das ursprüngliche sBA-Responsekriterium von „Woche 18“ auf „Woche 12“ geändert wurde (Protokollamendment 2), um dann erneut auf „Response zu Woche 12 oder 18“ in Protokollamendment 4, geändert zu werden. Bei letzterem waren bereits alle Testpersonen in die Studie ICONIC eingeschlossen. Eine Rationale für diese Änderung ist den Protokollen nicht zu entnehmen und eine ergebnisgesteuerte Auswertung nicht auszuschließen.

In der vorliegenden Nutzenbewertung werden stattdessen für die Ergebnisse der RW-Phase alle randomisierten Personen, die mindestens eine Dosis der Studienmedikation erhalten haben, berücksichtigt und für die einarmigen Ergebnisse die ITT-Population, die zudem der Sicherheitspopulation entspricht.

### Auswertungen im ItchRO

Die geplanten Auswertungen des ItchRO wurden im Laufe der Studie ICONIC mehrfach Änderungen unterzogen, auch nach Entblindung der Studie, weshalb es sich teilweise um post hoc bestimmte Auswertungen handelt. Eine ergebnisgesteuerte Berichterstattung ist nicht auszuschließen. Für die Änderungen konnten jeweils keine Begründungen identifiziert werden. Im Wesentlichen liegen 3 unterschiedliche Auswertungsmetriken für die Pruritus-Schwere (Item 1) vor:

- Kontinuierliche Auswertungen in wöchentlichen bzw. 4-wöchentlichen Durchschnittswerten.
- Responderanalysen der Erhebungen am Morgen ( $\geq 1$ ,  $\geq 1,25$ ,  $\geq 1,33$  und  $\geq 1,5$  Punkte).
- Kategoriale Auswertungen für die Anzahl und den Anteil der Tageswerte für die Erhebung am Morgen mit einem Wert von  $\leq 1$  Punkt.

In der vorliegenden Nutzenbewertung wird insbesondere die Reduktion des Pruritus-Werts von  $\leq 1$  Punkt als relevant erachtet, da dies keinem Pruritus oder nur einem geringen Pruritus entspricht. Responderanalysen auf Patientenebene liegen hierfür jedoch nicht vor und in den Ergebnissen wird deskriptiv die Anzahl und der Anteil der Tageswerte mit einem Wert von  $\leq 1$  Punkt dargestellt. Diese liegen jedoch ausschließlich für die Erhebung am Morgen vor.

Responderanalysen für eine Reduktion um  $\geq 1$ ,  $\geq 1,25$ ,  $\geq 1,33$  und  $\geq 1,5$  Punkte liegen ebenfalls nur für die Erhebung am Morgen vor und ausschließlich für den ItchRO(Obs). Zudem sind keine getrennten Auswertungen in der RW-Phase ersichtlich, sondern jeweils nur zu Respondern zu Woche 18 und zu Woche 48. Diese werden in den Ergebnissen bei den unkontrollierten Daten über den gesamten Studienverlauf dargestellt, allerdings nur für eine Reduktion um  $\geq 1$  Punkt (Verbesserung), da dieses Responderkriterium als plausibel für eine klinisch relevante Veränderung erachtet wird. Ergebnisse für eine Verschlechterung liegen nicht vor.

Somit liegen für die Responderanalysen zum einen nur selektive Auswertungen des ItchRO vor. Zum anderen konnten keine adäquaten Responderanalysen für die RW-Phase identifiziert werden, weshalb in der Nutzenbewertung auch die kontinuierlichen Auswertungen der wöchentlichen Durchschnittswerte berücksichtigt werden. Eine Bewertung der klinischen Relevanz der Mittelwertdifferenzen zwischen dem Maralixibat-Arm und dem Placebo-Arm erfolgte nicht. Ergebnisse zur Frequenz des Kratzens (Item 3 des ItchRO(Obs)) werden in der Ergebnisdarstellung nicht berücksichtigt, da sie als ein Teilaspekt der Pruritus-Schwere (Item 1) aufgefasst werden. Vor dem Hintergrund potentieller Carry-over-Effekte werden auch die 4-wöchentlichen Durchschnittswerte nicht dargestellt, da dieser Zeitraum der gesamten Dauer der RW-Phase entspricht.

## 2.5 Verzerrungspotential auf Studien- und Endpunktebene

Die Studie ICONIC liefert zum einen Daten zu einer 4-wöchigen RW-Phase, in der die Wirksamkeit und Sicherheit einer Weiterbehandlung mit Maralixibat in der Erhaltungsdosis im Vergleich zum Absetzen der Maralixibat-Behandlung auf ein Placebo bei denjenigen Personen untersucht wurde, die sich zu Woche 18 noch in der Studie befanden. Zum anderen liegen einarmige Daten der gesamten Studienpopulation für die Veränderung gegenüber Baseline vor. Die unkontrollierten Ergebnisse bis Woche 18 und nach der RW-Phase ab Woche 23 bis Woche 48 sind generell mit einem hohen Verzerrungspotential verbunden. Für die RW-Phase (Effekt des Absetzens von Maralixibat im Vergleich zur Weiterbehandlung) ist die Bewertung des Verzerrungspotentials auf Studien- und Endpunktebene nachfolgend dargestellt. Da es sich bei der Studie MRX-801 um eine Studie ohne Kontrollgruppe handelt, wird von einem hohen Verzerrungspotential auf Studien- und Endpunktebene ausgegangen.

Tabelle 10: Verzerrungspotential der RW-Phase der Studie ICONIC

| Studie | Adäquate Erzeugung der Randomisierungssequenz | Verdeckung der Gruppenzuteilung | Verblindung Patientin/Patient | Verblindung Behandlungsperson | Hinweise auf ergebnisgesteuerte Berichterstattung | Sonstige, das Verzerrungspotential beeinflussende Punkte | Verzerrungspotential auf Studienebene |
|--------|---|---------------------------------|-------------------------------|-------------------------------|---|--|---------------------------------------|
| ICONIC | Ja  | Unklar                          | Ja                            | Ja                            | Ja <sup>1)</sup>                                  | Ja <sup>2)</sup>   | Hoch                                  |

<sup>1)</sup> Die statistischen Auswertungen und das Studienprotokoll wurden mehrfach überarbeitet und der finale SAP erst nach Entblindung erstellt, sodass teilweise von post hoc geplanten Analysen ausgegangen werden muss.

<sup>2)</sup> Aufgrund fehlender Auswasch-Periode sind Carry-over-Effekte zu beachten.

Abkürzungen: SAP: Statistischer Analyseplan.

Das Verzerrungspotential auf Studienebene für die RW-Phase wird als hoch eingeschätzt. Das Verzerrungspotential der Endpunkte „Todesfälle“, „ItchRO“ und „PedsQL“ wird in Tabelle 11 dargestellt. Für die patientenberichteten Endpunkte „Klinische Kratzskala“, „Körperliche Entwicklung“ sowie „Unerwünschte Ereignisse“ liegen keine validen Ergebnisse für die RW-Phase vor (siehe Kapitel 3.3 und 3.5), sodass die Bewertung des Verzerrungspotentials entfällt und für die einarmigen Studienphasen von einem hohen Verzerrungspotential auszugehen ist.

**Tabelle 11: Verzerrungspotential der in der Nutzenbewertung berücksichtigten Endpunkte der RW-Phase der Studie ICONIC**

| Endpunkt                | Verblindung adäquat | ITT adäquat      | Hinweise auf ergebnisgesteuerte Berichterstattung | Sonstige, das Verzerrungspotential beeinflussende Punkte | Verzerrungspotential |
|-------------------------|---------------------|------------------|---|--|----------------------|
| Anzahl der Todesfälle   | Ja                  | Ja <sup>1)</sup> | Nein  | Nein   | Niedrig              |
| Tagebuch (ItchRO)       |                     |                  |   |  |                      |
| • Betreuerversion (Obs) | Ja                  | Ja <sup>1)</sup> | Ja <sup>7)</sup>                                  | Ja <sup>3)4)</sup>                                       | Hoch                 |
| • Kinderversion (Pt)    | Ja                  | Ja <sup>1)</sup> | Ja <sup>7)</sup>                                  | Ja <sup>2)3)4)</sup>                                     | Hoch                 |
| PedsQL-Fatigue          |                     |                  |   |  |                      |
| • Elternversion         | Ja                  | Ja <sup>1)</sup> | Nein  | Ja <sup>5)6)</sup>                                       | Hoch                 |
| • Kinderversion         | Ja                  | Ja <sup>1)</sup> | Nein  | Ja <sup>2)6)</sup>                                       | Hoch                 |
| PedsQL                  |                     |                  |   |  |                      |
| • Elternversion         | Ja                  | Ja <sup>1)</sup> | Nein  | Ja <sup>6)</sup>   | Hoch                 |
| • Kinderversion         | Ja                  | Ja <sup>1)</sup> | Nein  | Ja <sup>2)6)</sup>                                       | Hoch                 |

<sup>1)</sup> Von den 31 eingeschlossenen und mit Maralixibat behandelten Personen brachen 2 Personen vor Beginn der RW-Phase die Studie aufgrund von UE ab und wurden nicht randomisiert.

<sup>2)</sup> Da die Kinderversion erst bei Personen ab einschließlich 5 Jahren eingesetzt wurde, handelt es sich hierbei um eine Subpopulation (38 % der Maralixibat-Gruppe und 69 % der Placebo-Gruppe).

<sup>3)</sup> Aufgrund des Studiendesigns ist mit einem Carry-over-Effekt zu rechnen, der sich tendenziell zuungunsten von Maralixibat auswirkt.

<sup>4)</sup> Keine Angaben zum Anteil fehlender Werte in den Tagebucheinträgen.

<sup>5)</sup> Da für den PedsQL-Fatigue Versionen erst ab einem Alter von 2 Jahren existieren, handelt es sich hierbei um eine Subpopulation (77 % der Maralixibat-Gruppe und 88 % der Placebo-Gruppe).

<sup>6)</sup> Der Bezugszeitraum des Instruments von 4 Wochen entspricht der gesamten Dauer der RW-Phase. Gerade in der ersten Zeit nach Absetzen von Maralixibat sind Carry-over-Effekte möglich, die aufgrund des 4-wöchigen Bezugszeitraums erfasst werden.

<sup>7)</sup> Die geplanten Auswertungen des ItchRO wurden mehrfach überarbeitet und angepasst, teilweise erst nach Entblindung der Studie (siehe Kapitel 2.4).

Abkürzungen: ItchRO: Itch Reported Outcome; ITT: Intention-to-Treat; PedsQL: Pediatric Quality of Life Inventory; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; RW: Randomisierte Absetzphase; UE: Unerwünschtes Ereignis.

## 2.6 Indirekter Vergleich

Neben der Interventionsstudie ICONIC legt der pU im Nutzendossier ergänzend eine weitere Studie vor, GALA-MRX-ALGS, eine Untersuchung zur Langzeitbehandlung mit Maralixibat bei Personen mit ALGS im Vergleich zu einer externen Kontrollkohorte aus der GALA-Studie.

Dazu wurden mit Maralixibat behandelte Personen aus der pivotalen Studie ICONIC sowie aus 2 weiteren Studien (LUM001-301/303 und LUM001-302/305) verglichen mit unbehandelten Kontrollen aus der Studie Global ALagille Alliance (GALA), einer internationalen klinischen Multicenter-Datenbank zu ALGS. Diese Studie wurde im Verlauf des europäischen Zulassungsprozesses vorgelegt.

GALA-MRX-ALGS wird als „historischer Langzeit-Vergleich“ beschrieben. Es handelt sich um einen indirekten Vergleich ohne Brückenkompator.

Für den indirekten Vergleich wurden Auswertungen für die Endpunkte „Zeit des ereignisfreien Überlebens mit den Ereignissen Lebertransplantation, chirurgische Gallenableitung, Leberdekompensation und Tod jeglicher Ursache“ (primärer Endpunkt) sowie „Zeit des lebertransplantationsfreien Überlebens mit den Ereignissen Lebertransplantation und Tod jeglicher Ursache“ (post hoc definierter Endpunkt) vorgelegt.

### Beschreibung der Datenquellen

Zur Bildung der Studienpopulation, die den Wirkstoff Maralixibat erhalten hat (im Folgenden als Maralixibat-Kohorte bezeichnet), hat der pU Daten aller Teilnehmenden aus 3 Studien gepoolt: Der pivotalen Studie ICONIC sowie zwei Langzeitsicherheitsstudien (LUM001-301/305 und LUM001-302/303; siehe auch Abbildung 2). Bei den Studien LUM001-301/305 und LUM001-302/303 handelt es sich um Placebo-kontrollierte Studien mit anschließender einarmiger, Open-Label-Extensionsphase. In diesen beiden Studien wurde Maralixibat in Dosierungen verabreicht, welche nicht fachinformationskonform sind, aus diesem Grund wurden diese Studien nicht einzeln für die Nutzenbewertung eingereicht.

- Die pivotalen Studie ICONIC ist in den Tabellen 2 bis 4 der Nutzenbewertung beschrieben.
- In der Studie LUM001-301 erhielten Patientinnen und Patienten über 13 Wochen Maralixibat in einer Dosis von 280, 140 oder 70 µg/kg/Tag oder Placebo (n = 37), in der anschließenden einarmigen Extensionsstudie LUM001-305 (n = 34) erhielten die Patientinnen und Patienten bis zu 218 Wochen Maralixibat in der Dosis von 280 µg/kg/Tag.
- In der Studie LUM001-302 erhielten Patientinnen und Patienten über 13 Wochen Maralixibat in einer Dosis von 140 oder 70 µg/kg/Tag oder Placebo (n = 20), in der anschließenden einarmigen Extensionsstudie LUM001-303 erhielten die Patientinnen und Patienten bis zu 336 Wochen Maralixibat in einer Dosis von 280 µg/kg/Tag (n = 19) oder zweimal täglich 280 µg/kg/Tag (n = 5).

Alle Teilnehmenden waren zum Zeitpunkt des Studieneinschlusses zwischen 1 und 18 Jahre alt. Laut Integrated Clinical and Statistical Report [23] erhielten von ursprünglich 88 Personen 86 den Wirkstoff Maralixibat, von diesen entschieden sich 2 aus den Studien LUM001-301/302 nicht für die optionale Langzeit-Nachbeobachtung und wurden aus der gepoolten Maralixibat-Kohorte ausgeschlossen (siehe Abbildung 2). 84 Personen erhielten laut dem Report Maralixibat in den Extensionsstudien. Demgegenüber wurde allerdings im Studienbericht der Studie ICONIC dargelegt, dass 2 Personen die Studie in der Open-Label-Phase abgebrochen hatten, also in der optionalen Behandlungsverlängerung kein Maralixibat mehr erhalten hatten. Es bleibt unklar, warum diese beiden Personen in die gepoolte Maralixibat-Kohorte für den indirekten Vergleich aufgenommen worden sind.

Datenschnitt für die Maralixibat-Kohorte war der 31.03.2021. Zu diesem Zeitpunkt erhielten noch 38 Personen aus diesen abgeschlossenen Extensionsstudien weiterhin Maralixibat im Rahmen einer laufenden-Extensionsstudie (MRX-800).

Der integrierte Datensatz der Studien umfasst insgesamt Beobachtungszeiten bis 6 Jahre zur Untersuchung von Langzeiteffekten.



Abbildung 2: Übersicht über die Datenquellen der gepoolten Maralixibat-Kohorte für den indirekten Vergleich

Als Kontrollpopulation wurde eine selektierte Population der GALA-Studie verwendet, die nicht mit Maralixibat behandelt worden ist (im Folgenden als Kontrollkohorte bezeichnet). Die GALA-Studie ist als internationale „klinische Forschungsdatenbank“ beschrieben, in der klinische, genetische und Labordaten von Patientinnen und Patienten mit ALGS gesammelt werden. Es handelt sich um eine historische Kohorte, die aktuell Daten von > 1.400 Patientinnen und Patienten aus 67 pädiatrischen Zentren aus 35 Ländern umfasst [32,33]. Die Datenerfassung erfolgte retrospektiv zwischen Mai 2018 und Juli 2021. Die Studienzentren übermittelten die Daten über elektronische Case Report Forms in die zentrale Datenbank. Eingeschlossen wurden zwischen Januar 1997 und August 2019 geborene PatientInnen und Patienten mit einer genetisch bestätigten ALGS-Diagnose; vor 1997 geborene Kinder wurden eingeschlossen, wenn sie nachweislich eine pathogene oder wahrscheinlich pathogene JAG1- oder NOTCH2-Variante aufwiesen [33]. Die ALGS-Diagnose basierte, wie bei den Studien der Maralixibat-Kohorte, auf den Kriterien Familienvorgeschichte, pathogene Variante in JAG1 oder NOTCH2 und klinischen Charakteristika, allerdings ohne das Kriterium des Mangels an Gallengängen.

### Selektion der Kontrollen

Die Selektion der Kontrollen für den indirekten Vergleich mit der Maralixibat-Kohorte erfolgte in einem schrittweisen Verfahren auf Basis eines SAP [23]. Zunächst erfolgte eine Selektion der Kontrollen anhand der wichtigsten Ein- und Ausschlusskriterien der Maralixibat-Studien und anhand der in der GALA-Datenbank verfügbaren Variablen. Diese sind in der folgenden Tabelle gelistet.

*Tabelle 12: Schritt 1 der Selektion von Kontrollen: Identifikation der Personen und Visiten innerhalb der GALA-Studie*

| <b>Einschlusskriterien</b>   |
|--|
| <ul style="list-style-type: none"> <li>• Behandelt in Regionen, in denen die Maralixibat-Studien durchgeführt wurden: Nordamerika (USA, Kanada), Europa, Australien.</li> <li>• Alter bei Studienaufnahme: &gt; 12 Monate und &lt; 18 Jahre.</li> <li>• Cholestase, definiert durch eins oder mehrere der folgenden Kriterien: <ul style="list-style-type: none"> <li>○ Gesamtserumgallensäuren &gt; 3 x ULN</li> <li>○ Konjugiertes oder direktes Bilirubin &gt; 1 mg/dl</li> <li>○ Gesamtbilirubin &gt; 2 mg/dl</li> <li>○ GGT &gt; 3 x ULN</li> </ul> </li> </ul> |
| <b>Ausschlusskriterien</b>   |
| <ul style="list-style-type: none"> <li>• ALT &gt; 15 x ULN.</li> <li>• Klinisches Ereignis, definiert als Lebertransplantation, chirurgische Gallenableitung, Leberdekomensation oder Tod vor Aufnahme in die Studie.</li> <li>• Diagnose des hepatozellulären Karzinoms basierend auf der Histopathologie nach Biopsie.</li> <li>• Vor 1997 geboren.</li> <li>• Jedwede Teilnahme an einer Maralixibat-Studie.</li> <li>• Personen mit nur einer einzelnen Visite, für die keine Nachbeobachtungsinformation vorliegt.</li> </ul>                                   |

Abkürzungen: ALT: Alanin-Aminotransferase; GALA: Global ALagille Alliance; GGT: Gamma-Glutamyltransferase; ULN: Upper Limit of Normal.

Von den 1.438 Personen in der Datenbank wurden so 490 potentielle Kontrollen identifiziert.

Im Anschluss erfolgte die Auswahl einer dokumentierten Visite als Baseline-Visite (= der Beginn der Nachbeobachtung in Übereinstimmung mit den Maralixibat-Studien) mittels Maximum-Likelihood-Methoden. Diese Auswahl erfolgte gemäß SAP auf Basis aller verfügbaren Visiten im Vergleich zur Maralixibat-Kohorte sowie unter Berücksichtigung der Kovariaten „Alter bei Studieneinschluss“ und „Gesamtbilirubin“. Sensitivitätsanalysen wurden mit alternativen Baseline-Visiten durchgeführt.

469 Personen hatten alle Kovariaten im Datensatz, die für die Durchführung der Maximum-Likelihood-Auswahl benötigt wurden und bildeten die Kontrollkohorte. Von diesen verfügten 383 über eine Nachbeobachtungszeit von mindestens 3 Monaten, 308 erreichten eine Nachbeobachtungszeit von mindestens 6 Monaten und 205 erreichten eine Nachbeobachtungszeit von mehr als 12 Monaten.

#### Systematische Literaturrecherche und Auswahl von Confoundern

Eine systematische Suche nach Kontrollkohorten sowie eine systematische Confounder-Identifikation konnte den Unterlagen nicht entnommen werden.

#### Erhebung der klinischen Ereignisse

Die Erhebung der klinischen Ereignisse erfolgte in der Kontrollkohorte über die retrospektive Eingabe von Patientendaten durch die an der GALA-Studie beteiligten Studienzentren. Weitere Angaben (z. B. zur Visitensteuerung oder zur Erhebung von Todesfällen) waren aus den Unterlagen des pU und der Publikation zur GALA-Studie [33] nicht zu identifizieren. Klinische Ereignisse wurden von Patientinnen und Patienten nach ihrer letzten dokumentierten Visite nicht mehr erhoben. In den Maralixibat-Studien hingegen wurden klinische Ereignisse von Personen, die ihre Teilnahme beendeten, laut Studienprotokoll weiter erhoben, wobei in den Unterlagen des pU keine genaueren Angaben zum Prozedere identifiziert werden konnten. Lediglich zum

Endpunkt „Todesfall“ wird in Modul 4 beschrieben, dass die Information zum Tod aus öffentlich einsehbaren Datenquellen zusammengetragen wurde.

### Statistische Analysen

Zunächst sollte laut SAP die Ausgewogenheit zwischen der Maralixibat-Kohorte und der GALA-Kohorte zu Baseline hinsichtlich folgender Variablen bewertet werden:

- Alter
- Bilirubin (gesamt und direkt/konjugiert)
- Gamma-Glutamyltransferase (GGT)
- ALT

Für den Fall, dass die Verteilung dieser Variablen über die beiden Kohorten nicht ausgewogen ist, sollte eine Analyse mit Anpassung für Störfaktoren wiederholt werden. Als Kriterium für Vergleichbarkeit der beiden Kohorten wurde ein Cut-off für die standardisierte Mittelwertdifferenz von  $\leq 0,25$  gewählt. Inverse Probability of Treatment Weights (IPTW)-Methoden sollten in diesem Fall unter Verwendung der geschätzten Gewichte aus einer Propensity-Score-Methode angewendet werden.

#### **2.6.1 Zusammenfassende Bewertung der Methodik des indirekten Vergleichs**

Bei der in der Studie GALA-MRX-ALGS gebildeten Interventionsgruppe handelt es sich um 84 Patientinnen und Patienten mit ALGS aus 3 klinischen Studien, von denen über 60 % aus zwei Studien stammen, in denen Maralixibat in einer von der Fachinformation abweichenden, niedrigeren Dosis verabreicht wurde (maximal 280  $\mu\text{g}/\text{kg}/\text{Tag}$ ). Diese beiden Studien wurden als Einzelstudien auch vom pU explizit von der Nutzenbewertung ausgeschlossen. Voraussetzung für die Berücksichtigung klinischer Studien im Rahmen der Nutzenbewertung ist, dass der Wirkstoff in einer fachinformationskonformen Dosierung eingesetzt wurde. Die Intervention in dem vorgelegten indirekten Vergleich entspricht somit nicht der fachinformationskonformen Dosierung. Separate Auswertungen für die fachinformationskonform behandelte Population liegen nicht vor.

Zusätzlich zu der oben aufgeführten Limitation, die die Ablehnung des vorgelegten indirekten Vergleichs bedingt, ist bei den Ergebnissen auf folgende Auffälligkeiten und Einschränkungen hinzuweisen:

- Keine Begründung für die Auswahl der Kontrollkohorte bzw. Recherche nach weiteren, alternativen, natürlichen ALGS-Kohorten und keine systematische Recherche nach Confoundern

Aus den Unterlagen des pU ist nicht ersichtlich, dass eine systematische Literatursuche nach Registern oder Kohortenstudien zu ALGS durchgeführt worden ist, um beurteilen zu können, ob der Auswahl der Kontrollpopulation ein vollständiger Studienpool zugrunde gelegt wurde. Es wird lediglich aufgeführt, dass die GALA-Studie als Vergleichskohorte herangezogen wird. Eine Begründung für diese Auswahl liegt nicht vor. Bei einer orientierenden Recherche im Rahmen dieser Nutzenbewertung wurde eine weitere Studie identifiziert, die eine natürliche ALGS-Kohorte enthält, die Longitudinal Study of Genetic Causes of Intrahepatic Cholestasis (LOGIC). LOGIC ist im Studienregister clinicaltrials.org registriert [1] sowie im Repository des „National Institute of Diabetes and Digestive and Kidney Diseases“ [16]. Es handelt sich um eine laufende prospektive Studie, die im Jahr 2007 begonnen wurde. Die Studie enthält mindestens 252 Personen mit ALGS und nativer Leber (mindestens zu Baseline). In einer Publikation aus dem Jahr 2022 wurde gezeigt, dass es möglich war, aus dieser Kohorte eine Population von

59 Personen zu selektieren, die die Einschlusskriterien der Studie LUM001-301 erfüllten und der LUM001-301-Kohorte in wesentlichen klinischen Parametern ähnlich war [31]. Die Auswertung hat demonstriert, dass hier mindestens eine weitere natürliche Kohorte für einen indirekten Vergleich mit klinischen Maralixibat-Studien zur Verfügung steht. Ein Vorteil dieser Studie gegenüber der gewählten GALA-Studie zur Identifikation einer Kontrollpopulation besteht in der Verfügbarkeit von Daten zum Pruritus-Schweregrad, welcher in der Studie LOGIC über die Klinische Kratzskala operationalisiert wurde, ein Instrument, welches auch in den Maralixibat-Studien eingesetzt wurde. Vor dem Hintergrund dieser Kurzsrecherche ist demnach nicht davon auszugehen, dass das gewählte GALA-Register den bestmöglichen bzw. einzigen Pool an Kontrollpersonen darstellt.

- Keine Information zu Pruritus-Prävalenz und -Schweregrad in der Kontrollkohorte, sodass fraglich ist, wie viele Personen der Kontrollpopulation in das Anwendungsgebiet von Maralixibat fallen

Im Vergleich zu den Maralixibat-Studien wurden in der GALA-Studie wesentliche Ausschlusskriterien der Maralixibat-Studien, wie z. B. chronische Diarrhö, nicht berücksichtigt. Der bedeutendste Unterschied ist jedoch, dass für den Einschluss in eine Maralixibat-Studie ein moderater Pruritus bestehen musste, definiert als ein durchschnittlicher Tageswert von  $\geq 2$  im ItchRO-Fragebogen. Ob überhaupt ein Pruritus vorlag, wurde bei Einschluss in das GALA-Register nicht erhoben. Darüber hinaus wurden bei der Auswahl der Kontrollen aus dem Register, die für den indirekten Vergleich herangezogen wurden, für das Kriterium „Nachweis einer Cholestase“ andere Kriterien verwendet, als in den Maralixibat-Studien.

- Möglicher Selektionsbias aufgrund der Ein- und Ausschlusskriterien der Interventionsstudien von Maralixibat und nicht berichtete Kriterien für die Aufnahme der Kontrollpatientinnen und -patienten in die GALA-Studie, sodass die Vergleichbarkeit der behandelten und unbehandelten Kohorte fraglich ist

Bezüglich des Pruritus könnte aufgrund des erweiterten Einschlusskriteriums in den klinischen Studien (ItchRO  $\geq 2$  im Tagesmittelwert über 2 Wochen) eine Selektion zugunsten schwerer betroffener Fälle erfolgt sein. Auf der anderen Seite ist nicht bekannt, aus welchen Anlässen die Patientinnen und Patienten, die im GALA-Register verzeichnet sind, in den dokumentierenden Studienzentren vorstellig wurden. Hier ist eine Selektion möglich, z. B. durch Vorstellung aufgrund eines sich verschlechternden Zustands. Dadurch ist eine Selektion gesünderer und fitterer Personen in den klinischen Studien im Vergleich zur Kontrollpopulation nicht ausgeschlossen.

- Keine ausreichenden Angaben zu Patientencharakteristika (z. B. Krankheitsschwere) sowie Informationen zu Begleittherapien in der Kontrollpopulation zur Beurteilung der Vergleichbarkeit der Kohorten

Eine weitere Voraussetzung für die Durchführung eines indirekten Vergleichs ohne Brückenkomparator ist die Berücksichtigung aller relevanten Confounder als Adjustierungsvariablen, um die notwendige Strukturgleichheit zu erreichen. Im vorliegenden indirekten Vergleich wurden a priori 5 Kovariaten im SAP spezifiziert: Alter, Bilirubin (gesamt und direkt/konjugiert), GGT und ALT. Aus den Unterlagen des pU geht nicht hervor, wie diese Kovariaten identifiziert wurden. Es wird daher nicht davon ausgegangen, dass einer systematischen Herangehensweise zur Identifikation der Confounder gefolgt wurde. Eine vom pU in den Studienunterlagen als bedeutsam hervorgehobene klinische Variable – Pruritus-Schwere – war in der GALA-Studie nicht verfügbar und der vom pU als primärer Endpunkt in den klinischen Studien gewählte sBA-Wert (als ein Maß für Therapieerfolg) war nur bei weniger als 25 % der Kontrollkohorte aus der GALA-Studie verfügbar. Insgesamt ist unklar, ob die beiden Kohorten hinreichend vergleichbar sind.

- Selektive Endpunkt-Auswahl  
Eine weitere wesentliche Voraussetzung für die Berücksichtigung eines indirekten Vergleichs ist, dass dieser auf die interessierende Fragestellung insgesamt ausgerichtet ist und nicht nur selektiv auf z. B. einzelne Endpunkte. Der hier vorliegende Vergleich beinhaltet keine Analyse des Endpunkts „Pruritus“. Da das Anwendungsgebiet von Maralixibat die „Behandlung des Pruritus bei ALGS“ darstellt, ist hier eine weitere wesentliche Voraussetzung für einen indirekten Vergleich nicht erfüllt.
- Differentieller Drop-out und fehlende Daten bei über 50 % der Kontrollkohorte nach einem Jahr  
Die Erhebung der klinischen Ereignisse ist nur unzureichend beschrieben und erfolgte in den beiden Kohorten unterschiedlich. Bei Personen, die aus den klinischen Studien ausschieden, wurden klinische Ereignisse weiterhin erfasst, wobei unklar ist, mit welcher Methodik hier vorgegangen wurde. In der Kontrollkohorte wurden klinische Ereignisse nach der letzten Visite im klinischen Zentrum nicht nachverfolgt. Diese Praxis führte zu einer hohen und differentiellen Anzahl an Zensierungen zwischen Kontroll- und Maralixibat-Kohorte: In der Kontrollkohorte waren nach einem Jahr nur noch Endpunktdaten von 44 % der Kohorte verfügbar, im Vergleich zu 96 % in der Maralixibat-Kohorte. Nach 4 Jahren Follow-up waren auch für die Maralixibat-Kohorte nur noch Daten von 67 % der Patientinnen und Patienten verfügbar.

### 3 Ergebnisse der eingeschlossenen Studien

#### 3.1 Studiencharakteristika und Studienmedikation

Für einen Einschluss in die Studie ICONIC wurden 36 Personen gescreent, wovon 31 letztendlich eingeschlossen und behandelt wurden. In die Studie MRX-801 wurden alle 8 gescreenten Kleinkinder auch in die Studie eingeschlossen und haben Maralixibat erhalten.

Von den 31 Personen, die in die Studie ICONIC aufgenommen wurden, gingen 29 (94 %) in die RW-Phase über und 2 brachen zuvor aufgrund von UE ab. Nach Randomisierung erhielten 13 Personen weiterhin Maralixibat und 16 stattdessen Placebo für 4 Wochen. Alle 29 randomisierten Personen schlossen die RW-Phase ab und gingen in die langfristige einarmige Behandlungsphase bis Woche 48 über. Abgesehen von einem Studienabbruch aufgrund eines UE wurde die Kernphase bis Woche 48 von allen randomisierten Personen beendet. 23 Personen entschieden sich anschließend für eine Fortsetzung der Behandlung gemäß Protokollamendment 3 in der langfristigen Behandlungsverlängerung nach Woche 48, 5 stimmten dem nicht zu. Protokollamendment 4, welches eine nicht fachinformationskonforme Dosiserhöhung auf insgesamt 800 µg/kg/Tag ab Woche 100 vorsah, stimmten 4 Personen nicht zu. In der vorliegenden Nutzenbewertung werden ausschließlich Wirksamkeitsdaten bis einschließlich Woche 48 präsentiert, dem Ende der Kernstudie. Der Anteil der Studienteilnehmenden, für den nach Woche 48 Wirksamkeitsdaten berichtet sind, liegt für alle Wirksamkeitsendpunkten deutlich unter 70 %.

In die Studie MRX-801 wurden bis zum Interim-Datenschnitt vom 04.05.2022 8 Kleinkinder eingeschlossen und mit Maralixibat behandelt. 7 haben die 12-wöchige Kernstudienphase beendet und 6 gingen in die langfristigen Behandlungsphase ab Woche 13 über, in der sich zum Interim-Datenschnitt noch 5 Kinder befanden. Ein Kind brach in der Kernstudie aufgrund des Entzugs der Einverständniserklärung ab, und ein Kind nahm nicht an der langfristigen Behandlungsphase Teil (erhielt Maralixibat kommerziell). Die mediane Behandlungsdauer liegt bei 137 Tagen.

Tabelle 13: Allgemeine Angaben; Studien ICONIC und MRX-801

| Allgemeine Angaben                                       | ICONIC<br>Gesamt<br>n (%) | ICONIC<br>Maralixibat <sup>1)</sup><br>n (%) | ICONIC<br>Placebo <sup>1)</sup><br>n (%) | MRX-801<br>n (%) <sup>2)</sup> |
|--|---------------------------|--|--|--------------------------------|
| ITT-Population <sup>3)</sup>                             | 31 (100)                  | -  | -  | 8 (100)                        |
| Sicherheitspopulation                                    | 31 (100)                  | -  | -  | 8 (100)                        |
| RW-Population  | -                         | 13 (100)                                     | 16 (100)                                 | -                              |
| Abgeschlossene Behandlung                                | 14 (45) <sup>4)</sup>     | 13 (100)                                     | 16 (100)                                 | 1 (13) <sup>5)</sup>           |
| Abbruch Einnahme der Studienmedikation, n (%)            | 16 (52) <sup>6)</sup>     | 0  | 0  | 1 (13)                         |
| Aufgrund von:  |                           |  |  |                                |
| UE   | 5 (16)                    | -  | -  | -                              |
| Entscheidung Ärztin/Arzt                                 | 1 (3)                     | -  | -  | -                              |
| Entzug der Einverständniserklärung                       | 1 (3)                     | -  | -  | 1 (13)                         |
| Protokollamendment 3 nicht zugestimmt                    | 5 (16)                    | -  | -  | -                              |
| Protokollamendment 4 nicht zugestimmt                    | 4 (13)                    | -  | -  | -                              |
| Übergang in  |                           |  |  |                                |
| RW-Phase   | 29 (94) <sup>7)</sup>     | 13 (100)                                     | 16 (100)                                 | -                              |
| Langfristige Behandlungsphase                            | 29 (94)                   | -  | -  | 6 (75)                         |
| Optionale Behandlungsverlängerung                        | 23 (74) <sup>8)</sup>     | -  | -  | -                              |
| Langfristige optionale Behandlungsverlängerung           | 16 (52)                   | -  | -  | -                              |
| Beginn einer Behandlung 2x täglich                       | 15 (48)                   | -  | -  | -                              |
| Mittlere Behandlungsdauer <sup>9)</sup> , Tage (SD)      | 1.030 (668)               | 30 (4)                                       | 29 (3)                                   | 149 (50)                       |
| Mediane Behandlungsdauer <sup>9)</sup> , Tage (min; max) | 714 (36; 1.796)           | 28 (25; 40)                                  | 28 (25; 35)                              | 137 (101; 250)                 |
| Mediane Beobachtungsdauer                                | k. A.                     | k. A.  | k. A.                                    | k. A.                          |

- <sup>1)</sup> Angaben beziehen sich auf die 4-wöchige RW-Phase.
- <sup>2)</sup> Zum Interim-Datenschnitt vom 04.05.2022.
- <sup>3)</sup> Definition, siehe Kapitel 2.4.
- <sup>4)</sup> Als abgeschlossene Behandlung war definiert, dass sich die Person zum Zeitpunkt des finalen Datenschnitts noch in der Studie befand.
- <sup>5)</sup> Bis zum Datenschnitt wurde die langfristige Behandlungsverlängerung von 1 Person abgeschlossen (Übergang in das „Expanded Access Program“). 5 Personen befinden sich weiterhin unter Behandlung.
- <sup>6)</sup> Therapieabbrüche, die während der Studienphase mit 2 tägl. Dosierungen auftraten, sind nicht enthalten (N = 1).
- <sup>7)</sup> Von den 31 eingeschlossenen Personen brachen 2 die Studie aufgrund von UE in der offenen Behandlungsphase bis Woche 18 ab.
- <sup>8)</sup> Von 29 Personen stimmten 5 dem neuen Protokollamendment 3 nicht zu, 1 Person brach die Studie zuvor aufgrund von UE ab.
- <sup>9)</sup> Die Angaben zur Behandlungsdauer der Studie ICONIC schließen auch Dosierungen von MRX > 400 µg/kg/Tag ein. Informationen zur Behandlungsdauer von MRX ≤ 400 µg/kg/Tag liegen nicht vor.

Abkürzungen: ITT: Intention-to-Treat; k. A.: keine Angabe; MRX: Maralixibat; RW: Randomisierte Absetzphase; SD: Standardabweichung; UE: Unerwünschtes Ereignis.

Die Patientinnen und Patienten in der Studie ICONIC waren im Median 5 Jahre alt, wobei 16 in Altersgruppen ab 5 Jahre fielen (siehe Tabelle 14). Es waren etwas mehr männliche Personen eingeschlossen und sowohl bei der Größe als auch beim Gewicht zeigten sich deutlich negative Abweichungen von den Normwerten. Alle Studienteilnehmenden waren maximal 50 kg schwer. Außerdem ist bei allen Personen eine Mutation im JAGGED1 sowie eine Cholestase diagnostiziert und bei fast allen wurden in der Vergangenheit bereits Therapien gegen Juckreiz verabreicht. Von Xanthomen, die mit einer Beeinträchtigung einhergingen (Werte 3 und 4), waren nur 16 % der Studienteilnehmenden betroffen. Keine Angaben liegen zu Begleiterkrankungen vor.

Für die Studie MRX-801 liegen nur wenige Charakteristika zu Baseline vor. Die Kleinkinder waren im Mittel 6,5 Monate alt und fast alle männlich. Bereits hier zeigen sich deutlich negative Abweichungen bei den Parametern der körperlichen Entwicklung. Informationen hinsichtlich Diagnosekriterien, betreuerberichtetem Juckreiz (ItchRO(Obs)) sowie Herkunft der Kinder liegen nicht vor.

Tabelle 14: Charakterisierung der Studienpopulationen zu Baseline der Studien ICONIC und Studie MRX-801

| Charakterisierung der Studienpopulationen | Studie ICONIC       |                                     |                                 | Studie MRX-801       |
|---|---------------------|-------------------------------------|---------------------------------|----------------------|
|   | Gesamt<br>N = 31    | Maralixibat <sup>1)</sup><br>N = 13 | Placebo <sup>1)</sup><br>N = 16 | Maralixibat<br>N = 8 |
| <i>Alter (Jahre)</i>                      |                     |                                     |                                 | <i>(Monate)</i>      |
| MW (SD)                                   | 5,4 (4,25)          | 5,5 (5,03)                          | 5,8 (3,75)                      | 6,5 (2,51)           |
| Median (min; max)                         | 5,0 (1; 15)         | 4,0 (1; 15)                         | 5,0 (1; 14)                     | 7,0 (2; 10)          |
| <i>Altersgruppe (Jahre), n (%)</i>        |                     |                                     |                                 |                      |
| < 2 Jahre                                 | 6 (19,4)            | 3 (23,1)                            | 2 (12,5)                        | 8 (100)              |
| 2–4 Jahre                                 | 9 (29,0)            | 5 (38,5)                            | 3 (18,8)                        | -                    |
| 5–8 Jahre                                 | 9 (29,0)            | 2 (15,4)                            | 7 (43,8)                        | -                    |
| 9–12 Jahre                                | 4 (12,9)            | 1 (7,7)                             | 3 (18,8)                        | -                    |
| 13–18 Jahre                               | 3 (9,7)             | 2 (15,4)                            | 1 (6,3)                         | -                    |
| <i>Geschlecht, n (%)</i>                  |                     |                                     |                                 |                      |
| männlich                                  | 19 (61,3)           | 9 (69,2)                            | 10 (62,5)                       | 7 (87,5)             |
| weiblich                                  | 12 (38,7)           | 4 (30,8)                            | 6 (37,5)                        | 1 (12,5)             |
| <i>Land, n (%)</i>                        |                     |                                     |                                 |                      |
| Australien                                | 9 (29,0)            | 5 (38,5)                            | 4 (25,0)                        | k. A.                |
| Belgien                                   | 5 (16,1)            | 1 (7,7)                             | 2 (12,5)                        |                      |
| Frankreich                                | 9 (29,0)            | 3 (23,1)                            | 6 (37,5)                        |                      |
| Spanien                                   | 3 (9,7)             | 2 (15,4)                            | 1 (6,3)                         |                      |
| Polen                                     | 2 (6,5)             | 0                                   | 2 (12,5)                        |                      |
| Großbritannien                            | 3 (9,7)             | 2 (15,4)                            | 1 (6,3)                         |                      |
| <i>Größe (z-Score)</i>                    |                     |                                     |                                 |                      |
| MW (SD)                                   | -1,67 (1,34)        | -1,54 (1,26)                        | -1,84 (1,48)                    | -2,64 (1,59)         |
| Median (min; max)                         | -1,58 (-5,62; 0,42) | -1,67 (-4,22; 0,42)                 | -1,54 (-5,62; 0,13)             | -2,66 (-4,6; 0,5)    |
| <i>Gewicht (z-Score)</i>                  |                     |                                     |                                 |                      |
| MW (SD)                                   | -1,70 (1,18)        | -1,70 (1,33)                        | -1,60 (0,98)                    | -2,53 (1,55)         |
| Median (min; max)                         | -1,63 (-4,67; 0,11) | -1,67 (4,67; 0,11)                  | -1,61 (-3,37; 0,08)             | -2,27 (-5,3; -0,0)   |

| Charakterisierung der Studienpopulationen  | Studie ICONIC   |   |  | Studie MRX-801       |
|--|---|---|--|----------------------|
|  | Gesamt<br>N = 31  | Maralixibat <sup>1)</sup><br>N = 13   | Placebo <sup>1)</sup><br>N = 16  | Maralixibat<br>N = 8 |
| <i>Zeit seit ALGS-Diagnose (Monate)</i><br>MW (SD)<br>Median (min; max)  | 66,2 (50,38)<br>58,0 (5; 179)   | 64,5 (57,75)<br>44,0 (5; 179)   | 73,2 (45,50)<br>60,5 (13; 170)   | k. A.                |
| <i>Familiäre ALGS-Vorgeschichte, n (%)<sup>2)</sup></i><br>ja<br>nein<br>unbekannt   | 8 (25,8)<br>22 (71,0)<br>1 (3,2)  | 1 (7,7)<br>12 (92,3)<br>0   | 7 (43,8)<br>8 (50,0)<br>1 (6,3)  | k. A.                |
| <i>Mangel an Gallengängen, n (%)<sup>2)</sup></i><br>ja<br>nein<br>unbekannt   | 18 (58,1)<br>4 (12,9)<br>9 (29,0)   | 4 (30,8)<br>3 (23,1)<br>6 (46,2)  | 12 (75,0)<br>1 (6,3)<br>3 (18,8)   | k. A.                |
| <i>Vorhandensein einer Mutation, n (%)<sup>2)</sup></i><br>JAGGED1   | 31 (100)  | 13 (100)  | 16 (100)   | k. A.                |
| <i>Weitere klinische Merkmale des ALGS, n (%)<sup>2)3)</sup></i><br>Chronische Cholestase<br>Erkrankungen des Herzens<br>Nierenanomalien<br>Vaskuläre Anomalien<br>Skelettale Anomalien<br>Okuläre Anomalien<br>Charakteristische Gesichtsmerkmale | 31 (100)<br>29 (93,5)<br>12 (38,7)<br>5 (16,1)<br>17 (54,8)<br>17 (54,8)<br>29 (93,5) | 13 (100)<br>12 (92,3)<br>4 (30,8)<br>1 (7,7)<br>7 (53,8)<br>7 (53,8)<br>12 (92,3) | 16 (100)<br>15 (93,8)<br>8 (50,0)<br>3 (18,8)<br>9 (56,3)<br>8 (50,0)<br>15 (93,8) | k. A.                |
| Begleiterkrankungen  | k. A.   | k. A.   | k. A.  | k. A.                |
| <i>Vorthherapie(n) für Juckreiz zu Studienbeginn, n (%)</i><br>Ja  | 29 (93,5)   | 12 (92,3)   | 15 (93,8)  | k. A.                |
| <i>ItchRO(Obs) – Wöchentlicher Durchschnitt des maximalen Tageswerts</i><br>MW (SD)<br>Median (min; max)   | 3,1 (0,5)<br>3,1 (2,0; 4,0)   | 3,1 (0,4)<br>3,3 (2,43; 3,86)   | 3,1 (0,5)<br>3,0 (2,0; 4,0)  | k. A.                |

| Charakterisierung der Studienpopulationen          | Studie ICONIC       |                                     |                                 | Studie MRX-801       |
|--|---------------------|-------------------------------------|---------------------------------|----------------------|
|  | Gesamt<br>N = 31    | Maralixibat <sup>1)</sup><br>N = 13 | Placebo <sup>1)</sup><br>N = 16 | Maralixibat<br>N = 8 |
| <i>Klinische Xanthom-Skala, n (%)<sup>4)</sup></i> |                     |                                     |                                 |                      |
| 0  | 17 (54,8)           | 6 (46,2)                            | 9 (56,3)                        | k. A.                |
| 1  | 8 (25,8)            | 4 (30,8)                            | 4 (25,0)                        |                      |
| 2  | 1 (3,2)             | 1 (7,7)                             | 0                               |                      |
| 3  | 3 (9,7)             | 1 (7,7)                             | 2 (12,5)                        |                      |
| 4  | 2 (6,5)             | 1 (7,7)                             | 1 (6,3)                         |                      |
| <i>Serumgallensäure (µmol/l)</i>                   |                     |                                     |                                 |                      |
| MW (SD)  | 283,4 (210,6)       | 318,0 (233,7)                       | 249,6 (196,8)                   | 347,6 (275,1)        |
| Median (min; max)                                  | 275,6 (20,2; 748,5) | 335,4 (30,9; 748,5)                 | 195,8 (20,2; 583,4)             | 259,9 (136,2; 993,0) |
| <i>Gesamtbilirubin (mg/dl)</i>                     |                     |                                     |                                 |                      |
| MW (SD)  | 6,09 (5,78)         | 6,52 (6,57)                         | 4,83 (4,27)                     | 161,12 (104,87)      |
| Median (min; max)                                  | 4,60 (0,4; 20,5)    | 4,60 (0,5; 20,5)                    | 2,85 (0,4; 13,4)                | 158,76 (14,6; 337,3) |
| <i>Direktes Bilirubin (mg/dl)</i>                  |                     |                                     |                                 |                      |
| MW (SD)  | 4,57 (3,67)         | 4,69 (3,80)                         | 4,04 (3,59)                     | 124,55 (68,88)       |
| Median (min; max)                                  | 4,10 (0,2; 10,0)    | 4,10 (0,2; 10,0)                    | 2,70 (0,2; 10,0)                | 118,0 (8,5; 222,2)   |

<sup>1)</sup> Die Randomisierung erfolgte am Ende der offenen Behandlungsphase nach Woche 18. Die dargestellten Charakteristika beziehen sich auf die Baseline der Gesamtstudie.

<sup>2)</sup> Diagnostische Kriterium der ALGS – siehe Anhang.

<sup>3)</sup> Personen, für die mehr als ein klinisches ALGS-Merkmal berichtet ist, werden in jeder angegebenen Kategorie gezählt.

<sup>4)</sup> Die klinische Xanthom-Skala ist eine 5-Punkte-Skala (0 = kein; 1 = minimal; 2 = moderat; 3 = entstellend; 4 = behindernd).

Abkürzungen: ALGS: Alagille-Syndrom; ItchRO(Obs): Itch Reported Outcome (Observer); k. A.: keine Angabe; MW: Mittelwert; SD: Standardabweichung;

## Protokollverletzungen

In der Studie ICONIC ereigneten sich 5 relevante Protokollverletzungen, jeweils 2 Dosierungsfehler und Abweichungen von Ein- und Ausschlusskriterien. Bei einer Person des Placebo-Arms wurde die Visite zu Woche 22 aufgrund eines UE 3 Monate verspätet durchgeführt. Der pU präsentiert hierfür Sensitivitätsanalysen im ItchRO unter Ausschluss der betreffenden Person. Insgesamt ergeben sich jedoch anhand der in der Studie ICONIC dokumentierten Protokollverletzungen keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

Für die Studie MRX-801 sind keine Protokollverletzungen angegeben.

## Begleitmedikation

Im Interim-Report von Studie MRX-801 sind keine Begleitmedikationen berichtet; alle Personen der Studie ICONIC nahmen mindestens eine Begleitmedikation ein. Begleitmedikationen zur Behandlung des Pruritus in der Studie ICONIC sind in Tabelle 15 für den Zeitraum bis Woche 22 gelistet (offene Behandlungsphase bis Woche 18 und RW-Phase von Woche 19 bis Woche 22). Der Anteil der Studienteilnehmenden, die von Woche 23 bis Woche 48 Begleitmedikationen eingenommen haben, entspricht weitestgehend dem Anteil in der Behandlungsphase bis Woche 22. Neben diesen Medikationen gegen Pruritus erhielten die Patientinnen und Patienten in der Studie häufig (mehr als 20 %) Analgetika (ca. 60 %, insbesondere Paracetamol), Vitamine (> 90 %), Mittel bei säurebedingten Erkrankungen (> 35%, insbesondere Natriumbikarbonat) und Antihämorrhagika (ca. 80 %).

*Tabelle 15: Begleitmedikationen zur Behandlung des Pruritus bei mehr als 10 % der Studienpopulation nach Substanz; Studie ICONIC*

| <b>Studie ICONIC<br/>Begleitmedikationen<sup>1)</sup> zur Behandlung des Pruritus<sup>2)</sup></b> | <b>Gesamt<sup>3)</sup><br/>N = 31<br/>n (%)</b> | <b>Maralixibat<sup>4)</sup><br/>N = 13<br/>n (%)</b> | <b>Placebo<sup>4)</sup><br/>N = 16<br/>n (%)</b> |
|--|---|--|--|
| Jegliche Begleitmedikation zu Behandlung des Pruritus  | 28 (90)   | 12 (92)  | 14 (88)  |
| Rifampicin   | 23 (74)   | 10 (77)  | 11 (69)  |
| Phenobarbital  | 4 (13)  | 3 (23)   | 0  |
| Antihistaminika <sup>5)</sup>  | 4 (13)  | 3 (23)   | 4 (25)   |
| Ursodesoxycholsäure  | 25 (81)   | 10 (77)  | 13 (81)  |

<sup>1)</sup> Als Begleitmedikation wurde jegliche Medikation gewertet, die ab dem Zeitpunkt der ersten Dosis Studienmedikation bis einschließlich 14 Tage nach Einnahme der letzten Studienmedikation eingenommen wurde. Regelhaft sollte die Dosierung innerhalb der ersten 22 Wochen stabil gehalten werden. Zusätzlich durften keine Medikationen gegen Pruritus in den ersten 22 Wochen neu initiiert werden. Lediglich bei Antihistaminika kam es in mind. 3 Fällen zu einer Initiierung in der RW-Phase. Allerdings sind hier die Prävalenzen der Antihistaminika in beiden Gruppen balanciert.

<sup>2)</sup> Präspezifiziert gemäß Interim-SAP.

<sup>3)</sup> Begleitmedikationen innerhalb der offenen Behandlungsphase bis Woche 18.

<sup>4)</sup> RW-Phase.

<sup>5)</sup> Darunter fallen unterschiedliche Präparate.

Abkürzungen: RW: Randomisierte Absetzphase; SAP: Statistischer Analyseplan.

### 3.2 Mortalität

In der Studie ICONIC sind bis zum finalen Datenschnitt vom 21.08.2020 keine Todesfälle aufgetreten, genau wie in der Studie MRX-801 zum Interim-Datenschnitt vom 04.05.2022.

### 3.3 Morbidität

#### Pruritus mittels ItchRO

Der ItchRO(Pt) sollte von allen Personen ab einschließlich 9 Jahren erhoben werden, konnte jedoch bereits ab 5 Jahren (ggf. unter Hilfestellung durch die betreuende Person) zum Einsatz kommen. Zu Baseline waren 16 Personen der Studienpopulation der Studie ICONIC mindestens 5 Jahre alt und kamen für einen Patientenbericht in Frage. Der Anteil der Personen mit auswertbaren Patiententagebüchern lag bei jeder Erhebung bis Woche 48 oberhalb von 70 %, sodass auch die patientenberichteten Ergebnisse für die Subpopulation der  $\geq 5$ -Jährigen in der Nutzenbewertung berücksichtigt werden können. Zu beachten ist jedoch, dass die Anzahl der Personen im Placebo-Arm fast doppelt so hoch war, verglichen mit dem Maralixibat-Arm, und bei dieser Subpopulation die Strukturgleichheit der beiden Gruppen fraglich ist.

Für den ItchRO liegen eine Vielzahl von Auswertungen in der Studie ICONIC vor. Dabei werden primär die Responderanalysen als relevant erachtet. Diese liegen jedoch ausschließlich für die Erhebungen am Morgen vor. Zusätzlich werden Gruppenunterschiede in der Veränderung der Mittelwerte berichtet.

#### RW-Phase

In Tabelle 16 ist der Anteil der Tage mit Werten  $\leq 1$  Punkt in der ItchRO-Erhebung am Morgen sowohl betreuerberichtet als auch patientenberichtet dargestellt. Angaben zur Anzahl fehlender Werte liegen nicht vor und es fehlt eine statistische Auswertung zwischen den Behandlungsgruppen. Von den 435 dokumentierten Erhebungen während der RW-Phase bei den 13 Personen aus dem Maralixibat-Arm sind 190 (44 %) mit Werten  $\leq 1$  Punkt dokumentiert. Im Placebo-Arm waren es lediglich 127 (28 %) von 455 Erhebungen. Auch bei der patientenberichteten Version waren im Interventionsarm anteilig mehr Werte  $\leq 1$  Punkt (58 %) als im Vergleichsarm (42 %). Auswertungen auf Ebene der Studienteilnehmenden (z. B. Anteil der Personen, die im wöchentlichen Durchschnitt einen Wert von  $\leq 1$  Punkt erreichen) liegen nicht vor.

**Tabelle 16: Anteil der Tage bzw. Personen mit einem Wert von  $\leq 1$  Punkt im ItchRO in der Erhebung am Morgen; Studie ICONIC, RW-Phase (Woche 19 bis Woche 22)**

| Studie ICONIC<br>ItchRO  | Maralixibat<br>N = 13 <sup>1)</sup> | Placebo<br>N = 16 <sup>1)</sup> | Maralixibat<br>vs. Placebo |
|--|-------------------------------------|---------------------------------|----------------------------|
| <b>Erhebung am Morgen, betreuerberichtet (ItchRO(Obs))</b>                     |                                     |                                 |                            |
| <i>Woche 19 bis 22</i>   |                                     |                                 |                            |
| Gesamtzahl der Tage (= im Beobachtungszeitraum)                                | k. A.                               | k. A.                           | k. A.                      |
| Gesamtzahl der Tage mit Erhebung, N <sup>2)</sup>                              | 435                                 | 455                             |                            |
| Tage mit $\leq 1$ Punkt, n (%)   | 190 (44)                            | 127 (28)                        |                            |
| Anteil Personen mit wöchentlichem Durchschnittswert $\leq 1$ Punkt zu Woche 22 | k. A.                               | k. A.                           | k. A.                      |
| <b>Erhebung am Morgen, patientenberichtet (ItchRO(Pt))</b>                     |                                     |                                 |                            |
| <i>Woche 19 bis 22</i>   |                                     |                                 |                            |
| Gesamtzahl der Tage (= im Beobachtungszeitraum)                                | k. A.                               | k. A.                           | k. A.                      |
| Gesamtzahl der Tage mit Erhebung, N <sup>2)</sup>                              | 233                                 | 258                             |                            |
| Tage mit $\leq 1$ Punkt, n (%)   | 136 (58)                            | 109 (42)                        |                            |
| Anteil Personen mit wöchentlichem Durchschnittswert $\leq 1$ Punkt zu Woche 22 | k. A.                               | k. A.                           | k. A.                      |

<sup>1)</sup> RW-Population.

<sup>2)</sup> Anzahl der Erhebungen in der Bezugsperiode. Die maximal mögliche Anzahl an Erhebungen entspricht der kumulierten Anzahl der Tage, in denen sich alle Personen im jeweiligen Studienzeitraum befanden.

Abkürzungen: ItchRO: Itch Reported Outcome; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); k. A.: keine Angabe; RW: Randomisierte Absetzphase.

In Tabelle 17 ist die mittlere Veränderung im ItchRO(Obs) während der RW-Phase von Woche 22 im Vergleich zu Woche 18 dargestellt. Waren die Mittelwerte der wöchentlichen Durchschnitte der Erhebung am Morgen (nächtlicher Pruritus) zu Baseline zwischen der Maralixibat- und der Placebo-Gruppe mit 2,88 und 2,93 Punkten noch balanciert, zeigte sich bei der Erhebung zu Woche 18, Zeitpunkt der Randomisierung und Beginn der RW-Phase, ein etwas niedrigerer Wert im Placebo-Arm. Zu Woche 22 stieg der Wert dann im Placebo-Arm wieder fast auf den Ausgangswert zu Baseline an, während er in der Maralixibat-Gruppe auf dem niedrigen Niveau verblieb. Bei einem Unterschied von -1,5 Punkten zwischen den Behandlungsgruppen ist der Effekt dabei statistisch signifikant. Konsistente, statistisch signifikante Effekte zeigen sich sowohl bei der Erhebung am Abend (Pruritus am Tag) als auch beim schwersten Tageswert, bestehend aus dem wöchentlichen Durchschnitt der höheren der beiden täglichen Erhebungen. Sofern fehlende Werte mittels LOCF imputiert wurden, ergaben sich ebenfalls vergleichbare Ergebnisse. Auch in einer weiteren Sensitivitätsanalyse mittels MMRM für die Erhebung am Morgen im ItchRO zeigte sich ein statistisch signifikanter Effekt zugunsten von Maralixibat.

Auf Basis der vorliegenden Daten kann die klinische Relevanz der gezeigten Gruppenunterschiede in den gemittelten Veränderungswerten jedoch nicht beurteilt werden. Gleiches gilt für die Veränderungen im patientenberichteten Pruritus (s. Tabelle 18). Hier waren die Unterschiede zwischen Placebo und Maralixibat noch größer als bei den betreuerberichteten Einschätzungen. Allerdings ist zu beachten, dass es sich bei der Auswertung um eine Subpopulation von lediglich 55 % der randomisierten Population (alle Kinder ab 5 Jahre) handelt und im Placebo-Arm mehr als doppelt so viele Kinder waren wie im Maralixibat-Arm.

Tabelle 17: Veränderung des Pruritus im ItchRO(Obs) in der RW-Phase; Studie ICONIC

| Studie ICONIC<br>ItchRO(Obs)   | Maralixibat<br>N = 13 <sup>1)</sup> | Placebo<br>N = 16 <sup>1)</sup> |
|--|-------------------------------------|---------------------------------|
| <b>Erhebung am Morgen</b>  |                                     |                                 |
| <i>Baseline</i><br>MW (SD)   | 2,88 (0,54)                         | 2,93 (0,56)                     |
| <i>Woche 18</i><br>MW (SD)   | 1,30 (0,86)                         | 1,13 (0,85)                     |
| Woche 22, n (%)<br>MW (SD)   | 12 (92)<br>1,38 (0,93)              | 16 (100)<br>2,84 (0,85)         |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE)               | 12 (92)<br>0,22 (0,23)              | 16 (100)<br>1,700 (0,20)        |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup>                  | -1,48 [-2,12; -0,84]<br>< 0,0001    |                                 |
| <b>Erhebung am Abend</b>   |                                     |                                 |
| <i>Baseline</i><br>MW (SD)   | 2,70 (0,69)                         | 2,92 (0,55)                     |
| <i>Woche 18</i><br>MW (SD)   | 1,28 (0,84)                         | 1,13 (0,93)                     |
| Woche 22, n (%)<br>MW (SD)   | 1,45 (1,02)                         | 2,87 (0,82)                     |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>2)</sup> | 13 (100)<br>0,21 (0,2155)           | 16 (100)<br>1,71 (0,19)         |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup>                  | -1,50 [-2,10, -0,90]<br>< 0,0001    |                                 |
| <b>Maximum aus Abend und Morgen<sup>3)</sup></b>                         |                                     |                                 |
| <i>Baseline</i><br>MW (SD)   | 3,14 (0,42)                         | 3,10 (0,52)                     |
| <i>Woche 18</i><br>MW (SD)   | 1,51 (0,87)                         | 1,28 (0,93)                     |
| Woche 22, n (%)<br>MW (SD)   | 1,66 (1,07)                         | 3,06 (0,81)                     |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE)               | 13 (100)<br>0,21 (0,23)             | 16 (100)<br>1,73 (0,21)         |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup>                  | -1,52 [-2,16; -0,88]<br>< 0,0001    |                                 |

<sup>1)</sup> RW-Population.

<sup>2)</sup> Gruppenvergleich der Veränderung zwischen Woche 18 und Woche 22. Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

<sup>3)</sup> Ursprünglich geplante Auswertung gemäß Studienprotokoll.

Abkürzungen: ANCOVA: Analysis of Covariance; ItchRO(Obs): Itch Reported Outcome (Observer); KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler.

**Tabelle 18: Veränderung des Pruritus im ItchRO(Pt) in der RW-Phase; Studie ICONIC, RW-  
Population  $\geq 5$  Jahre zu Baseline**

| <b>Studie ICONIC<br/>ItchRO(Pt)</b>                                      | <b>Maralixibat<br/>N = 5<sup>1)</sup></b> | <b>Placebo<br/>N = 11<sup>1)</sup></b> |
|--|---|--|
| <b>Erhebung am Morgen</b>  |   |  |
| Baseline, n (%)<br>MW (SD)   | 5 (100)<br>2,85 (0,62)                    | 9 (82)<br>2,93 (0,72)                  |
| Woche 18, n (%)<br>MW (SD)   | 5 (100)<br>0,77 (0,76)                    | 9 (82)<br>0,87 (0,88)                  |
| Woche 22, n (%)<br>MW (SD)   | 5 (100)<br>0,68 (0,46)                    | 9 (82)<br>2,67 (0,92)                  |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>2)</sup> | 5 (100)<br>-0,15 (0,37)                   | 9 (82)<br>1,84 (0,28)                  |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup>                  |   | -1,99 [-3,01; -0,97]<br>0,0013         |
| <b>Erhebung am Abend</b>   |   |  |
| Baseline, n (%)<br>MW (SD)   | 5 (100)<br>2,20 (0,68)                    | 9 (82)<br>2,76 (0,73)                  |
| Woche 18, n (%)<br>MW (SD)   | 5 (100)<br>0,63 (0,46)                    | 9 (82)<br>0,70 (0,78)                  |
| Woche 22, n (%)<br>MW (SD)   | 5 (100)<br>0,63 (0,52)                    | 9 (82)<br>2,63 (0,92)                  |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>2)</sup> | 5 (100)<br>-0,03 (0,37)                   | 9 (82)<br>1,95 (0,28)                  |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup>                  |   | -1,98 [-3,00; -0,96]<br>0,0014         |
| <b>Maximum aus Abend und Morgen<sup>3)</sup></b>                         |   |  |
| Baseline, n (%)<br>MW (SD)   | 5 (100)<br>2,91 (0,54)                    | 9 (82)<br>3,08 (0,64)                  |
| Woche 18, n (%)<br>MW (SD)   | 5 (100)<br>0,94 (0,69)                    | 9 (82)<br>0,94 (0,92)                  |
| Woche 22, n (%)<br>MW (SD)   | 5 (100)<br>0,80 (0,55)                    | 9 (82)<br>2,91 (0,89)                  |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>2)</sup> | 5 (100)<br>-0,14 (0,37)                   | 9 (82)<br>1,97 (0,28)                  |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup>                  |   | -2,11 [-3,12; -1,09]<br>0,0008         |

<sup>1)</sup> RW-Population  $\geq 5$  Jahre zu Baseline.

<sup>2)</sup> Gruppenvergleich der Veränderung zwischen Woche 18 und Woche 22. Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

<sup>3)</sup> Ursprünglich geplante Auswertung gemäß Studienprotokoll.

Abkürzungen: ANCOVA: Analysis of Covariance; ItchRO(Pt): Itch Reported Outcome (Patient); KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler.

Abbildung 3 zeigt graphisch den Verlauf der Veränderung der Mittelwerte im ItchRO(Obs) im Vergleich zu Baseline bis Woche 48, getrennt nach Studienarm für die erhaltene Therapie während der RW-Phase. In beiden Gruppen zeigt sich über Woche 2, 6, 12 bis Woche 18 (Maralixibat-Applikation einarmig für alle Studienteilnehmenden) eine stetige Verbesserung im nächtlichen Pruritus (Erhebung am Morgen). Zu Woche 22, während der RW-Phase, gleicht sich die Placebo-Gruppe auf das Ausgangsniveau zu Baseline an, während im Maralixibat-Arm nach einem temporären Anstieg zu Woche 19 weiterhin eine Reduktion zu sehen ist. Zu Woche 28 erfolgt dann eine erneute Annäherung der Placebo-Gruppe an die verringerten Werte der Maralixibat-Gruppe, nachdem Maralixibat nach der RW-Phase erneut initiiert wurde. Für den ItchRO(Pt) zeigen sich ähnliche Verläufe für die Mittelwerte (siehe Abbildung 4).

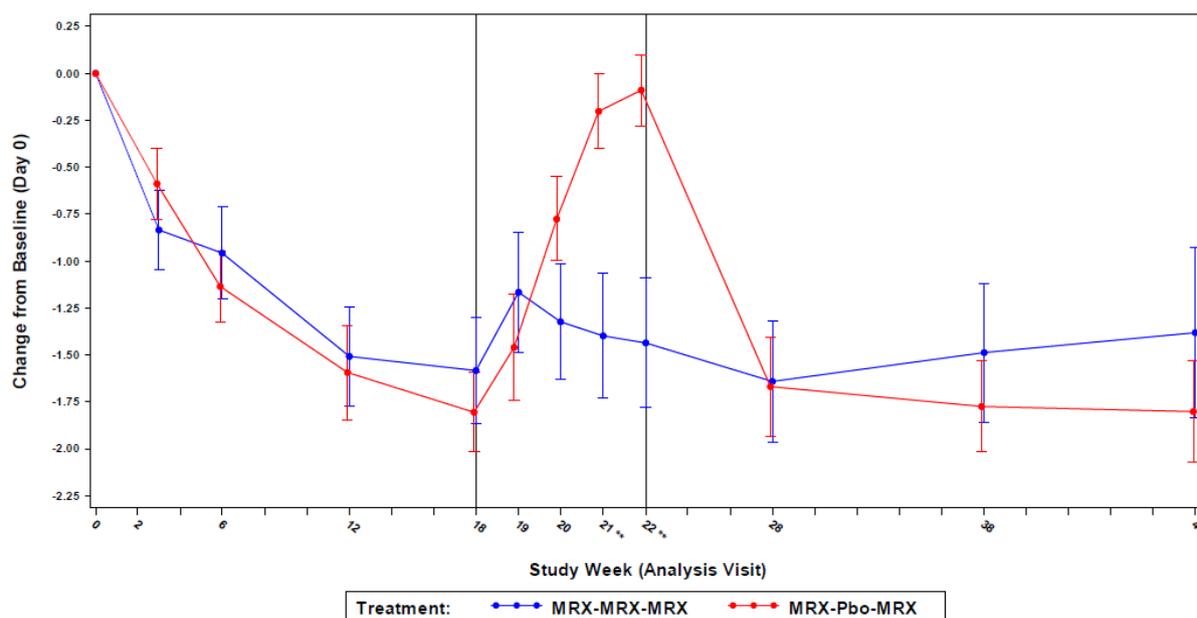


Abbildung 3: Mittelwert der Änderung des wöchentlichen morgendlichen ItchRO(Obs) gegenüber Baseline nach Behandlungsgruppe bis Woche 48 in der Studie ICONIC

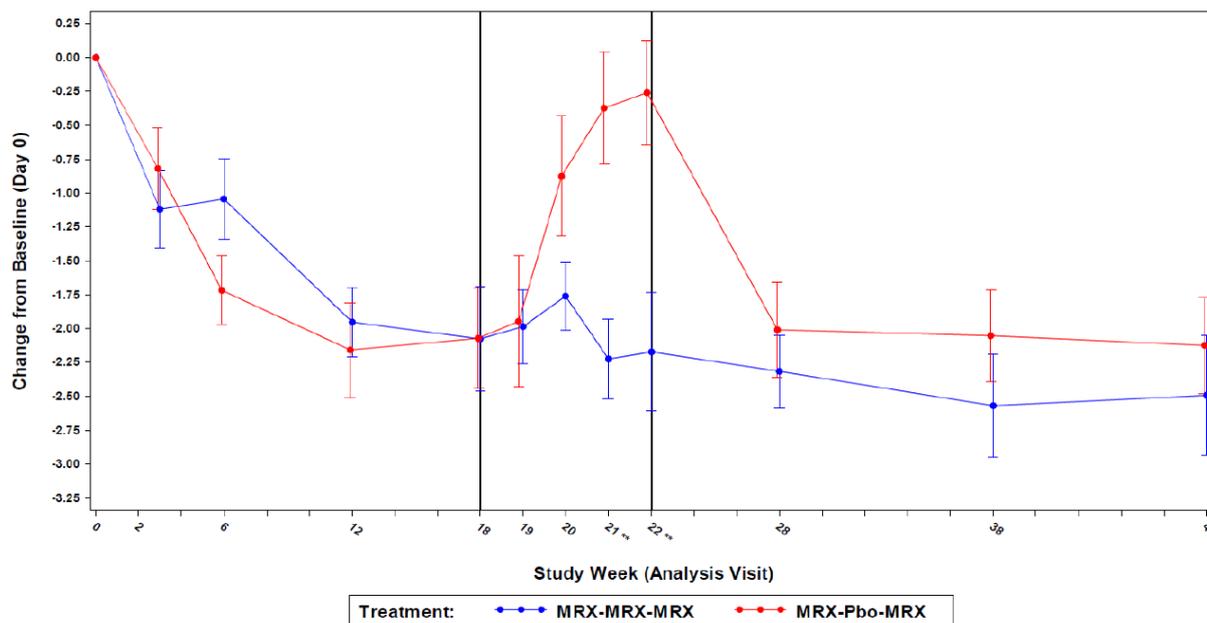


Abbildung 4: Mittelwert der Änderung des wöchentlichen morgendlichen ItchRO(Pt) gegenüber Baseline nach Behandlungsgruppe bis Woche 48 in der Studie ICONIC

Unkontrollierte Daten über den gesamten Studienverlauf (deskriptive Auswertung)

Da die RW-Phase der Studie ICONIC lediglich 4 Wochen dauerte und zudem die Ergebnisse ein hohes Verzerrungspotential aufweisen und somit die RW-Phase nicht ausreicht, um die Nachhaltigkeit des Therapieeffekts zu beurteilen, werden im Folgenden weitere einarmige Daten bis zur Erhebung an Woche 48 präsentiert. In Ergänzung zu Tabelle 16 und zur besseren Einordnung der Ergebnisse ist der Anteil der Tage mit Werten im ItchRO von  $\leq 1$  Punkt in der Erhebung am Morgen für die einzelnen Zeiträume der Studie ICONIC im Anhang in Tabelle 38 dargestellt; Ergebnisse auf Patientenebene liegen nicht vor. Sowohl in der betreuer- als auch der patientenberichteten Version des Tagebuchs steigt der Anteil der Erhebungen  $\leq 1$  Punkt vom Screening (6 %) zu Woche > 22 bis 48 kontinuierlich an auf 66 % im ItchRO(Obs) und 85 % im ItchRO(Pt). Die Anteile im ItchRO(Pt) waren dabei konstant höher als im ItchRO(Obs).

Zusätzlich zu der Anzahl der Erhebungen  $\leq 1$  Punkt im ItchRO sind in Tabelle 19 Responderanalysen für den ItchRO(Obs) für die Verbesserung im wöchentlichen Durchschnittswert der Erhebung am Morgen um mindestens 1 Punkt dargestellt. Zu Woche 48 wurde das Responderkriterium von 68 % der ITT-Population erreicht. Daten zum ItchRO(Pt) fehlen.

Tabelle 19: Responderanalysen zur Verbesserung im ItchRO in der Erhebung am Morgen; Studie ICONIC, ITT-Population

| Studie ICONIC ItchRO   | Gesamt ItchRO(Obs) N = 31 | Gesamt ItchRO(Pt) N = 16 <sup>1)</sup> |
|--|---------------------------|--|
| Woche 18<br>Verbesserung $\geq 1$ Punkt, n (%) <sup>2)</sup> | 21 (68)                   | k. A.                                  |
| Woche 48<br>Verbesserung $\geq 1$ Punkt, n (%)               | n = 29<br>21 (68)         | k. A.                                  |

<sup>1)</sup> Subpopulation der  $\geq 5$ -Jährigen in der Studie ICONIC.

<sup>2)</sup> Anteil bezogen auf die Gesamtpopulation; fehlende Werte wurden als Non-Responder gewertet.

Abkürzungen: ItchRO: Itch Reported Outcome; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); k. A.: keine Angabe.

Die Reduktion in den wöchentlichen Durchschnittswerten der Erhebung am Morgen hat sowohl im ItchRO(Obs) als auch im ItchRO(Pt) bis Woche 48 bestand (siehe Tabelle 20). Dies gilt auch für die weiteren durchgeführten Sensitivitätsanalysen der wöchentlichen Durchschnittswerte des ItchRO.

*Tabelle 20: Veränderung des Pruritus im ItchRO, Wert der Erhebung am Morgen; Studie ICONIC, ITT-Population*

| Studie ICONIC<br>ItchRO                      | Gesamt<br>ItchRO(Obs)<br>N = 31 | Gesamt<br>ItchRO(Pt)<br>N = 16 <sup>1)</sup> |
|--|---------------------------------|--|
| Baseline<br>MW (SD)                          | 2,91 (0,55)                     | 2,90 (0,66)                                  |
| Woche 48, n (%)<br>MW (SD)                   | 28 (90)<br>1,28 (1,14)          | 14 (88)<br>0,65 (0,76)                       |
| Woche 48 im Vergleich zu Baseline<br>MW (SD) | -1,62 (1,30)                    | -2,25 (1,01)                                 |

<sup>1)</sup> Subpopulation der  $\geq 5$ -Jährigen in der Studie ICONIC

<sup>2)</sup> Anzahl der beantworteten Fragen in der Bezugsperiode.

Abkürzungen: ItchRO: Itch Reported Outcome; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); MW: Mittelwert; SD: Standardabweichung.

### Studie MRX-801

Für Studie MRX-801 liegen keine Ergebnisse für den ItchRO(Obs) für den Interim-Datenschnitt vor.

### **Klinische Kratzskala**

#### RW-Phase

Responderanalysen, in denen ausschließlich die Veränderung in der RW-Phase bewertet wird, liegen nicht vor. Die vom pU präsentierte Auswertung Maralixibat vs. Placebo zu Woche 22 bezieht sich auf eine Veränderung gegenüber Baseline und nicht gegenüber Woche 18 (Beginn RW-Phase).

#### Unkontrollierte Daten über den gesamten Studienverlauf (deskriptive Auswertung)

Zu Baseline wiesen die Patientinnen und Patienten in der Studie ICONIC im Mittel relativ hohe Werte von 3,3 auf. Zu Woche 18, dem Beginn der RW-Phase, erreichten 21 Personen (68 %) eine Verbesserung um mindestens 1 Punkt. Zu Woche 48 erhöhte sich der Anteil der Responder weiter auf 77 %.

*Tabelle 21: Klinische Kratzskala, Responderanalyse; Studie ICONIC, ITT-Population*

| Studie ICONIC<br>Klinische Kratzskala <sup>1)</sup>                                      | Gesamt<br>N = 31   |
|--|--------------------|
| Baseline<br>MW (SD)  | 3,3 (0,9)          |
| Woche 18 im Vergleich zu Baseline, n (%)<br>Verbesserung um $\geq 1$ Punkt <sup>2)</sup> | 29 (93)<br>21 (68) |
| Woche 48 im Vergleich zu Baseline, n (%)<br>Verbesserung um $\geq 1$ Punkt <sup>2)</sup> | 28 (90)<br>24 (77) |

<sup>1)</sup> Skala von 0 bis 4, niedrigere Werte entsprechen einer besseren Symptomatik.

<sup>2)</sup> Bezogen auf die ITT-Population.

Abkürzungen: ITT: Intention-to-Treat.

### Studie MRX-801

In der Studie MRX-801 war die Variabilität im Pruritus gemäß der klinischen Kratzskala hoch, bei einem Mittelwert von 2,1 Punkten (SD: 1,3 Punkte). Von den 8 Kindern hatte 1 keinen Pruritus (Wert von 0), 2 hatten einen Wert von 1 (Reiben oder leichtes Kratzen, wenn keine Ablenkung erfolgt), 4 wiesen Werte zwischen 2 und 3 auf und bei 1 Kind waren Blutungen oder Narbenbildungen (Wert von 4) ersichtlich. Zu Woche 13, dem Ende der Erhaltungsphase, verbesserten sich von 7 Personen, für die Daten vorliegen, 4 um mindesten 1 Punkt; in 2 Fällen zeigte sich eine Verschlechterung um 1 bzw. 3 Punkte.

### **Körperliche Entwicklung**

#### RW-Phase

Auswertungen einer Veränderung der körperlichen Entwicklung in der RW-Phase, Woche 22 gegenüber Woche 18, liegen nicht vor. Die vom pU präsentierte Auswertung Maralixibat vs. Placebo zu Woche 22 bezieht sich auf eine Veränderung gegenüber Baseline und nicht gegenüber Woche 18 (Beginn der RW-Phase).

#### Unkontrollierte Daten über den gesamten Studienverlauf (deskriptive Auswertung)

Die Ergebnisse zur körperlichen Entwicklung anhand anthropometrischer Parameter sind in Tabelle 22 bis einschließlich Woche 48 dargestellt. In allen anthropometrischen Parametern zur Erfassung der körperlichen Entwicklung zeigten sich zu Studienbeginn deutlich negative Abweichungen von der Norm. Während sich der z-Score der Körpergröße zu Woche 48 im Vergleich zu Baseline in der Gesamtpopulation leicht um 0,178 verbesserte, zeigte sich im z-Score des Körpergewichts kaum ein Unterschied.

*Tabelle 22: Veränderung anthropometrischer Parameter; Studie ICONIC, ITT-Population*

| <b>Studie ICONIC<br/>Anthropometrische Parameter</b> | <b>Gesamt<br/>N = 31</b>  |
|--|---------------------------|
| <b>Körpergröße (z-Score)</b>                         |                           |
| <i>Baseline</i><br>MW (SD)                           | -1,668 (1,341)            |
| Woche 18, n (%)<br>MW (SD)                           | 29 (94)<br>-1,581 (1,292) |
| Veränderung von Baseline zu Woche 18<br>MW (SD)      | 0,124 (0,432)             |
| Woche 22, n (%)                                      | k. A.                     |
| Woche 48, n (%)<br>MW (SD)                           | 28 (90)<br>-1,436 (1,127) |
| Veränderung von Baseline zu Woche 48<br>MW (SD)      | 0,178 (0,501)             |

| Studie ICONIC<br>Anthropometrische Parameter    | Gesamt<br>N = 31          |
|---|---------------------------|
| <b>Körpergewicht (z-Score)</b>                  |                           |
| <i>Baseline</i><br>MW (SD)                      | -1,700 (1,184)            |
| Woche 18, n (%)<br>MW (SD)                      | 29 (94)<br>-1,624 (1,110) |
| Veränderung von Baseline zu Woche 18<br>MW (SD) | 0,019 (0,324)             |
| Woche 22, n (%)                                 | k. A.                     |
| Woche 48, n (%)<br>MW (SD)                      | 28 (90)<br>-1,517 (0,970) |
| Veränderung von Baseline zu Woche 48<br>MW (SD) | 0,018 (0,422)             |

Abkürzungen: ITT: Intention-to-Treat; k. A.: keine Angabe; MW: Mittelwert; SD: Standardabweichung.

### Studie MRX-801

Für die Studie MRX-801 liegen keine Ergebnisse zur körperlichen Entwicklung für den Interim-Datenschnitt vor.

### **PedsQL-Fatigue**

Für die elternberichtete Version der PedsQL-Fatigue existieren nur Versionen ab einem Alter von 2 Jahren. Dementsprechend liegen nur Daten für eine Subgruppe der Studie ICONIC vor, bestehend aus 25 Personen (81 %). Davon befinden sich in der RW-Phase 10 Personen (77 %) im Maralixibat-Arm und 14 (88 %) in der Placebo-Gruppe. Auch für die patientenberichtete Version des PedsQL-Fatigue (Beantwortung ab 5 Jahren) liegen analog zum ItchRO(Pt) nur Daten für eine Subgruppe (n=16) der Studie ICONIC vor. Die Rücklaufquote in beiden Subgruppen lag bis Woche 48 jeweils über 70%.

### RW-Phase

In den Tabellen 23 und 24 sind die Veränderungen in der PedsQL-Fatigue während der RW-Phase, Woche 22 im Vergleich zu Woche 18, dargestellt. Es zeigten sich weder in der Elternversion noch bei der von den Kindern selbst eingeschätzten Fatigue signifikante Gruppenunterschiede in der Veränderung zwischen Woche 18 und 22.

**Tabelle 23: Veränderung in der PedsQL-Fatigue (Parent) in der RW-Phase; Studie ICONIC, RW-Population  $\geq 2$  Jahre**

| <b>Studie ICONIC<br/>PedsQL-Fatigue (Parent)<sup>1)</sup></b>            | <b>Maralixibat<br/>N = 10<sup>2)</sup></b> | <b>Placebo<br/>N = 14<sup>2)</sup></b> |
|--|--|--|
| Baseline, n (%)<br>MW (SD)   | 9 (90)<br>60,34 (18,03)                    | 14 (100)<br>45,34 (23,84)              |
| Woche 18 <sup>3)</sup> , n (%)<br>MW (SD)                                | 9 (90)<br>74,85 (11,01)                    | 14 (100)<br>71,03 (14,88)              |
| Woche 22, n (%)<br>MW (SD)   | 9 (90)<br>67,64 (15,00)                    | 12 (86)<br>57,64 (23,39)               |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>2)</sup> | 9 (90)<br>-2,96 (6,05)                     | 16 (100)<br>-16,99 (5,24)              |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>4)</sup>                  | 14,03 [-2,78, 30,84]<br>0,0966             |  |

<sup>1)</sup> Werte von 1 bis 100, höhere Werte entsprechen einer besseren Symptomatik.

<sup>2)</sup> Für die elternberichtete Version der PedsQL-Fatigue existieren nur Versionen ab einem Alter von 2 Jahren. Die Subpopulation der  $\geq 2$ -Jährigen der Studie ICONIC besteht aus 10 Personen im Maralixibat-Arm und 14 in der Placebo-Gruppe.

<sup>3)</sup> Entspricht dem Baseline-Wert für die RW-Phase.

<sup>4)</sup> Gruppenvergleich der Veränderung zwischen Woche 18 und Woche 22. Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

Abkürzungen: ANCOVA: Analysis of Covariance; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler.

**Tabelle 24: Veränderung in der PedsQL-Fatigue (Child) in der RW-Phase; Studie ICONIC, RW-Population ab 5 Jahre**

| <b>Studie ICONIC<br/>PedsQL-Fatigue (Child)<sup>1)</sup></b>             | <b>Maralixibat<br/>N = 5<sup>2)</sup></b> | <b>Placebo<br/>N = 11<sup>2)</sup></b> |
|--|---|--|
| Baseline, n (%)<br>MW (SD)   | 4 (80)<br>60,42 (8,52)                    | 10 (91)<br>54,31 (15,79)               |
| Woche 18 <sup>3)</sup> , n (%)<br>MW (SD)                                | 5 (100)<br>61,94 (13,49)                  | 10 (91)<br>61,59 (10,48)               |
| Woche 22, n (%)<br>MW (SD)   | 5 (100)<br>63,33 (19,58)                  | 9 (82)<br>53,86 (15,37)                |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>2)</sup> | 5 (100)<br>1,05 (7,85)                    | 9 (82)<br>-8,53 (5,85)                 |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>4)</sup>                  | 9,57 [-11,97, 31,11]<br>0,3490            |  |

<sup>1)</sup> Werte von 1 bis 100, höhere Werte entsprechen einer besseren Symptomatik.

<sup>2)</sup> Randomisierte ITT-Population.

<sup>3)</sup> Entspricht dem Baseline-Wert für die RW-Phase.

<sup>4)</sup> Gruppenvergleich der Veränderung zwischen Woche 18 und Woche 22. Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

Abkürzungen: ANCOVA: Analysis of Covariance; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler.

### Unkontrollierte Daten über den gesamten Studienverlauf (deskriptive Auswertung)

Da die RW-Phase der Studie ICONIC lediglich 4 Wochen dauerte und somit nicht ausreicht, um die Nachhaltigkeit des Therapieeffekts zu beurteilen, werden in Tabelle 25 weitere einarmige Daten bis Woche 48 für die ITT-Population präsentiert. Sowohl in der Eltern- als auch in der Kinderversion zeigten sich zu Woche 18 und Woche 48 Verbesserungen, die in der Elternversion numerisch etwas ausgeprägter waren. Im Vergleich zu den Selbsteinschätzungen der Kinder war die Verbesserung der Fatigue in der Elternversion augenscheinlich stärker ausgeprägt. Ein direkter Vergleich der Ergebnisse der Elternversion vs. Kinderversion ist nicht möglich, da Selbsteinschätzungen altersbedingt nur für eine abweichende Teilpopulation vorliegen.

*Tabelle 25: Veränderung in der PedsQL-Fatigue; Studie ICONIC, ITT-Population ab 2 Jahren (Parent) und ab 5 Jahren (Child)*

| <b>Studie ICONIC<br/>PedsQL-Fatigue<sup>1)</sup></b> | <b>Gesamt<br/>PedsQL-Fatigue (Parent)<br/>N = 24<sup>2)</sup></b> | <b>Gesamt<br/>PedsQL-Fatigue (Child)<br/>N = 16<sup>3)</sup></b> |
|--|---|--|
| Baseline, n (%)<br>MW (SD)                           | 24 (100)<br>52,72 (23,30)   | 14 (88)<br>56,05 (14,06)   |
| Woche 18, n (%)<br>MW (SD)                           | 23 (96)<br>72,52 (13,36)  | 15 (94)<br>61,71 (11,07)   |
| Woche 18 im Vergleich zu Baseline, n (%)<br>MW (SD)  | 22 (97)<br>20,39 (25,89)  | 14 (88)<br>4,11 (17,69)  |
| Woche 48, n (%)<br>MW (SD)                           | 22 (97)<br>73,99 (15,46)  | 14 (88)<br>64,88 (19,17)   |
| Woche 48 im Vergleich zu Baseline, n (%)<br>MW (SD)  | 21 (88)<br>20,30 (24,87)  | 13 (81)<br>5,66 (13,91)  |

<sup>1)</sup> Werte von 1 bis 100, höhere Werte entsprechen einer besseren Symptomatik.

<sup>2)</sup> Für die elternberichtete Version der PedsQL-Fatigue existieren nur Versionen ab einem Alter von 2 Jahren. Die Subpopulation der  $\geq 2$ -Jährigen in der Studie ICONIC besteht aus 24 Personen.

<sup>3)</sup> Eine Beantwortung war ab einschließlich 5 Jahren möglich, wobei die Subpopulation der  $\geq 5$ -Jährigen in der Studie ICONIC aus 16 Personen bestand.

Abkürzungen: ITT: Intention-to-Treat; MW: Mittelwert; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; SD: Standardabweichung.

## **3.4 Lebensqualität**

Wie bereits bei der PedsQL-Fatigue beschrieben, waren 16 Personen der gesamten Studienpopulation der Studie ICONIC zu Baseline mindestens 5 Jahre alt und konnten den PedsQL Child beantworten.

### RW-Phase

In den Tabellen 26 und 27 sind die Veränderungen im PedsQL während der RW-Phase, Woche 22 im Vergleich zu Woche 18, dargestellt. Es zeigten sich weder bei der durch die Eltern eingeschätzten noch bei der von den Kindern selbst eingeschätzten gesundheitsbezogenen Lebensqualität statistisch signifikante Gruppenunterschiede in der Veränderung zwischen Woche 18 und Woche 22.

**Tabelle 26: Veränderung im PedsQL (Parent) in der RW-Phase; Studie ICONIC, RW-Population (Woche 18 bis Woche 22)**

| Studie ICONIC<br>PedsQL <sup>1)</sup> (Parent)                           | Maralixibat<br>N = 13          | Placebo<br>N = 16         |
|--|--------------------------------|---------------------------|
| Baseline<br>MW (SD)  | 64,79 (13,77)                  | 55,90 (17,80)             |
| Woche 18 <sup>2)</sup> , n (%)<br>MW (SD)                                | 12 (92)<br>74,80 (10,37)       | 16 (100)<br>68,24 (16,43) |
| Woche 22, n (%)<br>MW (SD)   | 13 (100)<br>66,89 (15,36)      | 16 (100)<br>59,87 (20,27) |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>3)</sup> | 12 (92)<br>-6,69 (4,51)        | 16 (100)<br>-9,03 (3,89)  |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>3)</sup>                  | 2,33 [-10,08; 14,75]<br>0,7018 |                           |

<sup>1)</sup> Werte von 1 bis 100, höhere Werte entsprechen einer besseren Lebensqualität.

<sup>2)</sup> Entspricht dem Baseline-Wert für die RW-Phase.

<sup>3)</sup> Gruppenvergleich der Veränderung zwischen Woche 18 und Woche 22. Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

Abkürzungen: ANCOVA: Analysis of Covariance; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PedsQL: Pediatric Quality of Life Inventory; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler.

**Tabelle 27: Veränderung im PedsQL (Child) in der RW-Phase; Studie ICONIC, RW-Population ab 5 Jahre (Woche 18 bis Woche 22)**

| Studie ICONIC<br>PedsQL <sup>1)</sup> (Child)                            | Maralixibat<br>N = 5          | Placebo<br>N = 11        |
|--|-------------------------------|--------------------------|
| Baseline, n (%)<br>MW (SD)   | 5 (100)<br>68,91 (14,70)      | 10 (91)<br>61,74 (11,09) |
| Woche 18 <sup>2)</sup> , n (%)<br>MW (SD)                                | 5 (100)<br>66,74 (11,07)      | 10 (91)<br>70,43 (10,10) |
| Woche 22, n (%)<br>MW (SD)   | 5 (100)<br>71,52 (14,75)      | 9 (82)<br>66,30 (11,64)  |
| Veränderung von Woche 18 bis Woche 22, n (%)<br>LS MW (SE) <sup>3)</sup> | 5 (100)<br>4,35 (4,46)        | 9 (82)<br>-3,50 (3,31)   |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>3)</sup>                  | 7,85 [-4,44; 20,13]<br>0,1874 |                          |

<sup>1)</sup> Werte von 1 bis 100, höhere Werte entsprechen einer besseren Lebensqualität.

<sup>2)</sup> Entspricht dem Baseline-Wert für die RW-Phase.

<sup>3)</sup> Gruppenvergleich der Veränderung zwischen Woche 18 und Woche 22. Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

Abkürzungen: ANCOVA: Analysis of Covariance; ITT: Intention-to-Treat; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PedsQL: Pediatric Quality of Life Inventory; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler

### Unkontrollierte Daten über den gesamten Studienverlauf (deskriptive Auswertung)

Sowohl in der Eltern- als auch in der Kinderversion zeigten sich zu Woche 18 Verbesserungen, die zu Woche 48 etwas geringer ausfielen. Im Vergleich zu den Selbsteinschätzungen der Kinder war die Verbesserung in der Lebensqualität in der Elternversion etwas stärker ausgeprägt. Allerdings ist auch hier ein direkter Vergleich der Ergebnisse der Eltern- vs. Kinderversion nicht möglich, da Selbsteinschätzungen altersbedingt nur von etwas mehr als 50 % der Patientinnen und Patienten vorlagen.

*Tabelle 28: Veränderung der Lebensqualität im PedsQL; Studie ICONIC, ITT-Population*

| <b>Studie ICONIC<br/>PedsQL</b>   | <b>Gesamt<br/>PedsQL (Parent)<br/>N = 31</b> | <b>Gesamt<br/>PedsQL (Child)<br/>N = 16<sup>1)</sup></b> |
|---|--|--|
| Baseline, n (%)<br>MW (SD)  | 31 (100)<br>61,10 (16,99)                    | 15 (94)<br>64,13 (12,37)                                 |
| Woche 18, n (%)<br>MW (SD)<br><i>Woche 18 im Vergleich zu Baseline</i><br>MW (SD) | 28 (90)<br>71,05 (14,31)<br>10,73 (16,24)    | 15 (94)<br>69,20 (10,19)<br>5,07 (13,16)                 |
| Woche 48, n (%)<br>MW (SD)<br><i>Woche 48 im Vergleich zu Baseline</i><br>MW (SD) | 27 (87)<br>68,34 (15,50)<br>8,94 (18,74)     | 14 (88)<br>69,88 (17,69)<br>4,35 (15,15)                 |

<sup>1)</sup> Eine Beantwortung war ab einschließlich 5 Jahren möglich, wobei die Subpopulation der  $\geq 5$ -jährigen in der Studie ICONIC aus 16 Personen bestand.

Abkürzungen: ITT: Intention-to-Treat; MW: Mittelwert; PedsQL: Pediatric Quality of Life Inventory; SD: Standardabweichung.

## 3.5 Sicherheit

### **Unerwünschte Ereignisse**

Wie bereits ausgeführt werden die Ergebnisse der UE innerhalb der RW-Phase der Studie ICONIC als nicht valide erachtet. Es werden ausschließlich die nicht-vergleichenden Ergebnisse der einarmigen Behandlungsphasen präsentiert. Für die vorliegende Nutzenbewertung sind prinzipiell alle UE in jeglicher Studienphase relevant, die sich bei einer Dosis von maximal 400 µg/kg/Tag Maralixibat ereignen. Aggregierte Ergebnisse für alle UE, die unter Maralixibat  $\leq 400$  µg/kg/Tag aufgetreten sind, liegen nicht vor. Die folgende Ergebnisdarstellung bezieht sich daher separat auf UE bei der Erhaltungsdosis von 400 µg/kg/Tag Maralixibat und auf UE, die sich bis zu einer Dosis von 400 µg/kg/Tag ereigneten, was in den Auftitrationsphase und bei Dosisreduktionen der Fall war. Dabei gehen auch UE ab Woche 48 in die Auswertungen ein. Angaben zur Beobachtungsdauer liegen nicht vor.

Für die Studie MRX-801 sind dagegen alle UE dargestellt, die unter Maralixibat auftraten, sowohl während der Auftitration als auch während der Erhaltungphase.

Alle Ergebnisse beziehen sich jeweils auf die Sicherheitspopulationen, wobei in Studie ICONIC der finale Datenschnitt vom 21.08.2020 und in der Studie MRX-801 ein Interim-Datenschnitt vom 04.05.2022 Berücksichtigung fand. Ausgenommen davon sind die UE  $< 400$  µg/kg/Tag Maralixibat in Studie ICONIC, für die lediglich Daten aus dem Interim-Report mit einem Datenschnitt vom 01.12.2019 zu entnehmen sind.

**Tabelle 29: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Behandlungsphase; Studien ICONIC und MRX-801**

| Zusammenfassung der UE<br><i>Personen mit mindestens einem ...</i> | ICONIC Gesamt   |   |   | MRX-801                       |
|--|---|---|---|-------------------------------|
|  | Maralixibat<br>< 400 µg/kg/Tag <sup>1)</sup><br>N = 31<br>n (%) | Maralixibat<br>≤ 400 µg/kg/Tag<br>N = 31<br>n (%) | Maralixibat<br>400 µg/kg/Tag<br>N = 31<br>n (%) | Maralixibat<br>N = 8<br>n (%) |
| UE   | 27 (87,1)   | k. A.   | 30 (96,8)                                       | 7 (88)                        |
| UE CTCAE-Grad ≥ 3 <sup>2)</sup>                                    | 1 (3,2)   |   | 9 (29,0)  | 4 (50)                        |
| SUE  | 0   |   | 9 (29,0)  | 4 (50)                        |
| UE, das zum Abbruch der Studienmedikation führte                   | 1 (3,2)   |   | 5 (16,1)  | 0                             |

<sup>1)</sup> Diese Dosiskategorie umfasst die UE unter den Dosierungen der Auftitrationsphase in Woche 1–6, allen Auftitrierungen nach möglichen Dosisunterbrechungen sowie entsprechenden Dosisreduzierungen. In den weiteren Tabellen ist diese Dosiskategorie aufgrund fehlender Angaben nicht aufgeführt.

<sup>2)</sup> Falls in den CTCAE keine Graduierung für ein bestimmtes UE verfügbar ist, erfolgt die Einstufung in „mild“, „moderat“ oder „schwer“ durch das Prüfpersonal.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; k. A.: keine Angabe; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

**Tabelle 30: UE mit Inzidenz ≥ 10 % nach MedDRA-Systemorganklassen und Preferred Terms; Studien ICONIC und MRX-801<sup>1)</sup>**

| UE mit Inzidenz ≥ 10 %<br><b>MedDRA-Systemorganklasse</b><br><i>Preferred Term</i> | ICONIC Gesamt                                     |   | MRX-801 <sup>2)</sup>         |
|--|---|---|-------------------------------|
|  | Maralixibat<br>≤ 400 µg/kg/Tag<br>N = 31<br>n (%) | Maralixibat<br>400 µg/kg/Tag<br>N = 31<br>n (%) | Maralixibat<br>N = 8<br>n (%) |
| <b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>                | k. A.   | <b>18 (58,1)</b>                                | <b>4 (50,0)</b>               |
| Pyrexie  |   | 15 (48,4)                                       | 3 (37,5)                      |
| <b>Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>                  |   | <b>16 (51,6)</b>                                | <b>3 (38,0)</b>               |
| Husten   |   | 11 (35,5)                                       | -                             |
| Schmerzen im Oropharynx  |   | 5 (16,1)  | -                             |
| Rhinorrhoe   |   | -   | 2 (25,0)                      |
| <b>Erkrankungen der Haut und des Unterhautgewebes</b>                              |   | <b>5 (16,1)</b>                                 | <b>2 (25,0)</b>               |
| Pruritus   |   | 4 (12,9)  | -                             |
| <b>Erkrankungen des Blutes und des Lymphsystems</b>                                |   | <b>4 (12,9)</b>                                 | -                             |
| <b>Erkrankungen des Gastrointestinaltrakts</b>                                     |   | <b>23 (74,2)</b>                                | <b>5 (62,5)</b>               |
| Abdominalschmerz   |   | 16 (51,6)                                       | 3 (37,5)                      |
| Diarrhö  |   | 11 (35,5)                                       | 3 (37,5)                      |
| Erbrechen  |   | 11 (35,5)                                       | -                             |
| Zahnen   |   | -   | 3 (37,5)                      |

| UE mit Inzidenz $\geq 10\%$<br><i>MedDRA-Systemorganklasse</i><br><i>Preferred Term</i> | ICONIC Gesamt   |  | MRX-801 <sup>2)</sup>         |
|---|---|--|-------------------------------|
|   | Maralixibat<br>$\leq 400 \mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>400 $\mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>N = 8<br>n (%) |
| <b>Erkrankungen des Nervensystems</b>   |   | <b>9 (29,0)</b>  | <b>2 (25,0)</b>               |
| Kopfschmerzen   |   | 6 (19,4)   | -                             |
| <b>Erkrankungen des Ohrs und des Labyrinths</b>   |   | <b>4 (12,9)</b>  | -                             |
| <b>Infektionen und parasitäre Erkrankungen</b>  |   | <b>25 (80,6)</b>   | <b>6 (75,0)</b>               |
| Nasopharyngitis   |   | 9 (29,0)   | 4 (50,0)                      |
| Infektion der oberen Atemwege   |   | 8 (25,8)   | -                             |
| Ohreninfektion  |   | 6 (19,4)   | -                             |
| Gastroenteritis   |   | 5 (16,1)   | -                             |
| Virusinfektion  |   | 6 (19,4)   | -                             |
| <b>Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen</b>                         |   | <b>4 (12,9)</b>  | -                             |
| <b>Untersuchungen</b>   |   | <b>5 (16,1)</b>  | -                             |
| <b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>               |   | <b>13 (41,9)</b>   | -                             |
| Sturz   |   | 5 (16,1)   | -                             |

<sup>1)</sup> Anteil Personen mit mind. einem UE je SOC und PT (Personen wurden nur einmal gezählt pro SOC und PT).

<sup>2)</sup> Für die Studie MRX-801 wurden nur UE nach SOC und PT dargestellt, die bei mind. 2 Personen aufgetreten sind.

Abkürzungen: -: UE mit Inzidenz  $< 10\%$  bzw. bei mind. 2 Personen; k. A.: keine Angabe; MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Term; SOC: Systemorganklasse; UE: Unerwünschtes Ereignis.

### Schwere unerwünschte Ereignisse

Schwere UE CTCAE-Grad 3 oder höher sind in Tabelle 31 aufgeführt. Keine Angaben liegen für UE vor, die in der Studie ICONIC insgesamt bis zu einer Dosis von einschließlich  $400 \mu\text{g/kg/Tag}$  aufgetreten sind.

Tabelle 31: Schwere UE (Grad 3 oder höher), die bei  $\geq 5\%$  der Personen aufgetreten sind, nach Systemorganklasse und Preferred Term; Studien ICONIC und MRX-801<sup>1)</sup>

| Schwere UE (Grad 3 oder höher), die bei $\geq 5\%$ der Personen aufgetreten sind<br><i>MedDRA-Systemorganklasse</i><br><i>Preferred Term</i> | ICONIC Gesamt   |  | MRX-801 <sup>2)</sup>         |
|--|---|--|-------------------------------|
|  | Maralixibat<br>$\leq 400 \mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>400 $\mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>N = 8<br>n (%) |
| <b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>  |   | -  | <b>2 (25)</b>                 |
| <b>Erkrankungen des Blutes und des Lymphsystems</b>  | k. A.   | <b>2 (6,5)</b>   | -                             |
| <b>Erkrankungen des Gastrointestinaltrakts</b>   |   | <b>2 (6,5)</b>   | <b>2 (25)</b>                 |
| Abdominalschmerz   |   | 2 (6,5)  | -                             |
| <b>Infektionen und parasitäre Erkrankungen</b>   |   | -  | <b>2 (25)</b>                 |

| Schwere UE (Grad 3 oder höher), die bei $\geq 5\%$ der Personen aufgetreten sind<br><i>MedDRA-Systemorganklasse</i><br><i>Preferred Term</i> | ICONIC Gesamt   |  | MRX-801 <sup>2)</sup>         |
|--|---|--|-------------------------------|
|  | Maralixibat<br>$\leq 400 \mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>400 $\mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>N = 8<br>n (%) |
| <b>Untersuchungen</b>  |   | <b>3 (9,7)</b>   | -                             |
| Alanin-Aminotransferase erhöht   |   | 2 (6,5)  | -                             |
| <b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>  |   | <b>2 (6,5)</b>   | -                             |

<sup>1)</sup> Anteil Personen mit mind. einem UE je SOC und PT (Personen wurden nur einmal gezählt pro SOC und PT).

<sup>2)</sup> Für die Studie MRX-801 wurden nur UE nach SOC und PT dargestellt, die bei mind. 2 Personen aufgetreten sind.

Abkürzungen: -: Schwere UE mit Inzidenz  $< 5\%$  bzw. bei mind. 2 Personen; k. A.: keine Angabe; MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Term; SOC: Systemorganklasse; UE: Unerwünschtes Ereignis.

### Schwerwiegende unerwünschte Ereignisse

SUE sind in Tabelle 32 aufgeführt. Keine Angaben liegen für SUE vor, die in der Studie ICONIC insgesamt bis zu einer Dosis von einschließlich  $400 \mu\text{g/kg/Tag}$  aufgetreten sind. Sowohl in der Studie ICONIC als auch in der Studie MRX-801 trat kein Preferred Term eines SUE bei mehr als 1 Person auf.

Tabelle 32: SUE, die bei  $\geq 5\%$  der Personen aufgetreten sind, nach Systemorganklasse und Preferred Term; Studien ICONIC und MRX-801, Sicherheitspopulation<sup>1)</sup>

| SUE, die bei $\geq 5\%$ der Personen aufgetreten sind<br><i>MedDRA-Systemorganklasse</i><br><i>Preferred Term</i> <sup>3)</sup> | ICONIC Gesamt   |  | MRX-801 <sup>2)</sup>         |
|---|---|--|-------------------------------|
|   | Maralixibat<br>$\leq 400 \mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>400 $\mu\text{g/kg/Tag}$<br>N = 31<br>n (%) | Maralixibat<br>N = 8<br>n (%) |
| <b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>   |   | -  | <b>2 (25)</b>                 |
| <b>Erkrankungen des Gastrointestinaltrakts</b>  | k. A.   | <b>3 (9,7)</b>   | -                             |
| <b>Infektionen und parasitäre Erkrankungen</b>  |   | <b>5 (16,1)</b>  | <b>2 (25)</b>                 |
| <b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>   |   | <b>2 (6,5)</b>   | -                             |

<sup>1)</sup> Anteil Personen mit mind. einem UE je SOC und PT (Personen wurden nur einmal gezählt pro SOC und PT).

<sup>2)</sup> Für die Studie MRX-801 wurden nur UE nach SOC und PT dargestellt, die bei mind. 2 Personen aufgetreten sind.

<sup>3)</sup> Es traten keine PT mit einer Inzidenz von  $\geq 5\%$  auf.

Abkürzungen: -: SUE mit Inzidenz  $< 5\%$  bzw. bei mind. 2 Personen; k. A.: keine Angabe; MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Term; SOC: Systemorganklasse; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

### UE von besonderem Interesse

UE von besonderem Interesse waren in den Studien ICONIC und MRX-801 unterschiedlich definiert. In der Studie MRX-801 ist bei keinem Kleinkind ein UE von besonderem Interesse aufgetreten. Für die Studie ICONIC sind die Ergebnisse in Tabelle 33 dargestellt. Ergebnisse für schwere und schwerwiegende UE von besonderem Interesse liegen nicht vor.

Tabelle 33: UE von besonderem Interesse; Studie ICONIC

| Studie ICONIC <sup>1)</sup><br>UE von besonderem Interesse <sup>2)</sup><br>(unabhängig vom Schweregrad) | Maralixibat<br>≤ 400 µg/kg/Tag<br>N = 31<br>n (%) | Maralixibat<br>400 µg/kg/Tag<br>N = 31<br>n (%) |
|--|---|---|
|  | Diarrhö <sup>3)</sup>                             | k. A.   |
| Ereignisse im Zusammenhang mit erhöhtem Bilirubin <sup>4)</sup>  | 1 (3,2)   |   |
| Ereignisse im Zusammenhang mit erhöhten Transaminasen <sup>5)</sup>                                      | 3 (9,7)   |   |
| Ereignisse im Zusammenhang mit Mängeln der fettlöslichen Vitamine <sup>6)</sup>                          | 6 (19,4)  |   |

<sup>1)</sup> Die Daten wurden den Unterlagen des pU zum Interim-Datenschnitt (01.12.2019) entnommen, da zum finalen Datenschnitt keine Angaben für diese UE in den Dosisklassen über den gesamten Studienzeitraum vorliegen.

<sup>2)</sup> In den Studienunterlagen des Dossiers wird der Begriff „Adverse Events of Special Interest“ bzw. „UE von speziellem Interesse“ verwendet.

<sup>3)</sup> Umfasst UE der PT „Diarrhö“ oder „Gastroenteritis“.

<sup>4)</sup> Definiert als PT „Bilirubin im Blut erhöht“.

<sup>5)</sup> Umfasst UE der PT „Alanin-Aminotransferase erhöht“ oder „Aspartat-Aminotransferase erhöht“.

<sup>6)</sup> Umfasst viele im Anhang des SAP gelistete, SOC-übergreifende PT.

Abkürzungen: k. A.: keine Angabe; PT: Preferred Term; pU: pharmazeutischer Unternehmer; SAP: Statistischer Analyseplan; SOC: Systemorganklasse; UE: Unerwünschtes Ereignis.

## 4 Diskussion der Methodik und Ergebnisse

### 4.1 Zulassungsstatus und Zulassungspopulation von Maralixibat

Maralixibat (Livmarli®) ist gemäß Fachinformation zugelassen zur Behandlung des cholestatischen Pruritus bei Patientinnen und Patienten mit Alagille-Syndrom (ALGS) ab dem Alter von 2 Monaten [6]. Für Maralixibat liegt eine Zulassung als Arzneimittel zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden vor. Die empfohlene Zieldosis (Erhaltungsdosis) liegt bei 380 µg/kg einmal täglich. Die Anfangsdosis beträgt 190 µg/kg einmal täglich und sollte nach einer Woche auf 380 µg/kg einmal täglich erhöht werden. Maralixibat 380 µg/kg/Tag als freie Base entspricht einer Dosis von Maralixibat-Chlorid von 400 µg/kg/Tag. Die Verabreichung erfolgt täglich am Morgen oral über eine Applikationspritze.

Die Zulassung basiert auf der pivotalen Studie ICONIC, in die Kinder und Jugendliche bis 18 Jahre mit ALGS, einer bestehenden Cholestase und Pruritus eingeschlossen waren. Zusätzlich wurde Evidenz aus der Studie MRX-801 herangezogen, die Säuglinge (< 12 Monate) mit cholestatischen Lebererkrankungen (PFIC oder ALGS) einschloss. Beide Studien wurden in der vorliegenden Nutzenbewertung berücksichtigt.

Sowohl in der Studie ICONIC als auch in der Studie MRX-801 war ein von der Fachinformation abweichendes Auftitrationsschema vorgesehen, bei dem Maralixibat über einen Zeitraum von 6 Wochen von 14 µg/kg/Tag auf 400 µg/kg/Tag eskaliert wurde. Gemäß European Public Assessment Report (EPAR) wurde u. a. aufgrund einer guten Verträglichkeit zur Bestimmung der fachinformationskonformen Anwendung die oben beschriebene vereinfachte Auftitration in einem Dosisschritt (190 µg/kg/Tag) in 1 Woche gewählt. Die Auswirkungen dieser Abweichung, insbesondere auf die Wirksamkeitsdaten der RW-Phase der Studie ICONIC, werden als gering eingestuft. Weiterhin war in der Studie ICONIC in der langfristigen Behandlungsverlängerung ab Woche 100 eine weitere Dosis Eskalation auf insgesamt 800 µg/kg/Tag vorgesehen, was nicht fachinformationskonform ist. In der vorliegenden Nutzenbewertung werden daher ausschließlich die Ergebnisse bis zu einer Dosis von 400 µg/kg/Tag berücksichtigt.

Hinsichtlich der Studienpopulation waren sowohl in der Studie ICONIC als auch in der Studie MRX-801 Personen mit Lebertransplantation oder einer chirurgischen Unterbrechung des enterohepatischen Kreislaufs von der Teilnahme ausgeschlossen. Der pU geht davon aus, dass Maralixibat nach Lebertransplantation nicht mehr eingesetzt wird, da der hepatologische Defekt inklusive der auftretenden Symptome als geheilt gilt. Daten zu Wirksamkeit und Sicherheit bei Personen mit Eingriffen in den enterohepatischen Kreislauf fehlen jedoch und es ist unklar, ob diese im selben Maß von einer Therapie mit Maralixibat profitieren. Zum Anteil dieser Personen in der Versorgung liegen keine Daten vor. Zudem hatte keine Person eine Mutation im NOTCH2-Signalweg, deren Prävalenz ist jedoch deutlich geringer im Vergleich zu der bei allen Studienteilnehmenden vorhandenen JAGGED1-Mutation [18]. In der Studie ICONIC beschränkte sich die Studienpopulation weiterhin auf Personen mit einem mindestens moderaten Pruritus (definiert als durchschnittlicher Tageswert > 2 im ItchRO) und einem Gewicht von maximal 50 kg zu Baseline. Eine Rationale für die Gewichtsbeschränkung liegt nicht vor und die Wirksamkeit von Maralixibat bei leichtem Pruritus ist unklar. Das Bestehen eines Pruritus zu Baseline war in Studie MRX-801 kein Einschlusskriterium.

Mangels ausreichender Daten zu Parametern der Cholestase wurde statt des ursprünglich angestrebten Anwendungsgebiets „Behandlung der Cholestase bei Personen mit ALGS“ eine Beschränkung auf die Behandlung des cholestatischen Pruritus bei Patientinnen und Patienten mit ALGS vorgenommen.

Maralixibat wurde unter „außergewöhnlichen Umständen“ zugelassen.

## 4.2 Design und Methodik der Studien

### Studie ICONIC

Die Nutzenbewertung stützt sich insbesondere auf die pivotale Studie ICONIC, einer multi-zentrischen, offenen Langzeitstudie mit einer doppelblinden, Placebo-kontrollierten, RW-Phase zur Untersuchung der Wirksamkeit und Sicherheit von Maralixibat bei Kindern und Jugendlichen mit ALGS. Die Studie gliedert sich in insgesamt 6 Phasen: Eine offene Dosisescalationsphase über 6 Wochen, in der Maralixibat auf bis zu 400 µg/kg/Tag auftitriert wurde; eine 12-wöchige, offene Behandlungsphase mit stabiler Dosierung; eine 4-wöchige randomisierte, doppelblinde, Placebo-kontrollierte Absetzphase (RW-Phase); eine 26-wöchige, langfristige Behandlungsphase mit stabilen Dosierungen bis zu 400 µg/kg/Tag – diese ersten 4 Phasen bis Woche 48 stellen die Kernstudie dar. Nach Woche 48 bestand die Möglichkeit an einer optionalen Behandlungsverlängerung über 52 Wochen bis Woche 100 teilzunehmen. Danach war eine ebenfalls optionale Teilnahme an einer langfristigen Behandlungsverlängerung möglich, wobei Maralixibat auf maximal 800 µg/kg/Tag (400 µg/kg zweimal täglich) erhöht werden konnte, abhängig von der Wirksamkeit und Sicherheit. Diese letzte Phase dauerte an, bis zur kommerziellen Verfügbarkeit des Medikaments, der Teilnahme an einer weiteren Maralixibat-Studie oder der Einstellung des Entwicklungsprogramms. In die Studie ICONIC wurden 31 Personen aus 9 Studienzentren in Europa und Australien eingeschlossen, davon keines in Deutschland. Die Studie wurde am 28.05.2020 abgeschlossen. Primärer Endpunkt war die Veränderung des Nüchtern-sBA-Levels zu Woche 22 gegenüber Woche 18 bei Personen, die zuvor auf die Maralixibat-Behandlung angesprochen haben. Weitere Endpunkte waren u. a. die Bewertung des Pruritus mittels Tagebuch und klinischer Kratzskala, die körperliche Entwicklung, Xanthome, Leberfunktion und PedsQL sowie unerwünschte Ereignisse.

Zum Ende der ersten offenen Behandlungsphase (Woche 0 bis Woche 18) wurden alle Personen, die sich zu diesem Zeitpunkt noch in der Studie befanden, im Verhältnis 1:1 in die Maralixibat- oder die Placebo-Gruppe randomisiert, stratifiziert nach dem Ansprechen im sBA-Wert (definiert als Reduktion von  $\geq 50\%$  im sBA-Wert zu Woche 12 oder Woche 18). In der anschließenden 4-wöchigen RW-Phase wurden die Patientinnen und Patienten der Maralixibat-Gruppe in konstanter Dosis weiter mit Maralixibat behandelt, während im Kontrollarm stattdessen ein Placebo verabreicht wurde. Im Vergleich zu vielen anderen Studien mit einem Withdrawal-Design wurde in der Studie ICONIC jedoch kein Vorgehen gewählt, in dem die Randomisierung ausschließlich auf Personen beschränkt wurde, die zuvor ein Ansprechen auf die Therapie gezeigt haben. Von 31 in die Studie ICONIC eingeschlossenen Kindern und Jugendlichen wurden 29 randomisiert, davon 13 in die Maralixibat-Gruppe und 16 in die Placebo-Gruppe. 2 Personen brachen die Studie bis Woche 18 ab. Die RW-Phase wurde von allen 29 Personen abgeschlossen. In der anschließenden einarmigen langfristigen Behandlungsphase bis Woche 48 wurden Personen der Placebo-Gruppe erneut in 6-wochen auf 400 µg/kg/Tag Maralixibat auftitriert. In der Maralixibat-Gruppe erfolgte eine Schein-Auftitrierung. Das vorgesehene Studienende der Studie ICONIC zu Woche 48 wurde in 2 Protokollamendments um eine optionale Behandlungsverlängerung zunächst auf 100 Wochen verlängert (Protokollamendment 3), dann um eine optionale langfristige Behandlungsverlängerung auf unbestimmte Zeit (Protokollamendment 4). Beim Übergang wurde das Studienmedikament bei insgesamt 13 Personen über einen längeren Zeitraum (d. h. mindestens 7 Wochen) abgesetzt, bis die Amendments implementiert wurden und eine Weiterbehandlung möglich war. Für die Weiterbehandlung nach Woche 48 war eine erneute Einverständniserklärung der Patientinnen und Patienten erforderlich, die jedoch von 5 Personen nicht erteilt wurde. Lediglich 23 Personen (73 %) traten die optionale Behandlungsverlängerung ab Woche 49 an.

Nach Woche 48 lagen nur noch Wirksamkeitsdaten für weniger als 70 % der Studienpopulation vor. Da es zudem bei einem relevanten Teil der Studienpopulation beim Übergang zu langfristigen Behandlungsunterbrechungen kam, werden für die vorliegende Nutzenbewertung ausschließlich die Ergebnisse bis Woche 48 für die Wirksamkeitsendpunkte dargestellt.

Eine Person (randomisiert in die Placebo-Gruppe) hat die RW-Phase abgeschlossen, unterbrach dann allerdings in Woche 23 die Behandlung für 3 Monate aufgrund eines UE. Die Visite zu Woche 22 fand nicht zum geplanten Zeitpunkt statt, sondern wurde nach der 3-monatigen Unterbrechung durchgeführt. In einer Reihe von Sensitivitätsanalysen wurde gezeigt, dass der Ausschluss dieser Person nicht zu einer Veränderung der Ergebnisse im ItchRO führten. Für den PedsQL wurden allerdings keine Sensitivitätsanalysen berichtet.

Die Kinder und Jugendlichen in der Studie ICONIC waren im Median 5 Jahre alt, wobei 16 älter als 5 Jahre waren. In die Studie waren etwas mehr männliche Probanden eingeschlossen (61 %) und sowohl bei der Größe als auch beim Gewicht zeigten sich deutlich negative Abweichungen von den Normwerten. Bei allen Patientinnen und Patienten ist eine Mutation im JAGGED1 sowie eine Cholestase diagnostiziert und bei fast allen wurden bereits Therapien gegen Juckreiz verabreicht (94 %). Von Xanthomen, die mit einer Beeinträchtigung einhergingen (Werte 3 und 4), waren nur 16 % der Studienteilnehmenden betroffen.

Eine erneute Charakterisierung der Studienpopulation zu Woche 18, dem Beginn der RW-Phase, wurde nicht vorgenommen. Die präsentierten Patientencharakteristika der Behandlungsgruppen Maralixibat und Placebo beziehen sich ebenfalls auf die Baseline der Studie ICONIC. Patientinnen und Patienten der Maralixibat-Gruppe waren zu Baseline etwas jünger (Median 4 vs. 5 Jahre), wiesen höhere Werte im sBA auf (Median 335 vs. 196  $\mu\text{mol/l}$ ) und hatten deutlich weniger häufig eine familiäre ALGS-Vorgeschichte (92 vs. 50 %) sowie einen Mangel an Gallengängen (31 vs. 75 %) auf. Die Gruppen waren jedoch hinsichtlich anthropometrischer Parameter, Geschlecht und Pruritus zu Baseline mittels ItchRO(Obs) weitestgehend vergleichbar. Aufgrund der teilweise deutlichen Unterschiede in einzelnen Patientencharakteristika bei jedoch sehr geringen Patientenzahlen, kann die Vergleichbarkeit der Behandlungsgruppen in der RW-Phase nicht abschließend bewertet werden.

Die Einnahme von Begleitmedikationen zur Behandlung des Pruritus war erlaubt, unter der Voraussetzung einer stabil gehaltenen Dosis in den ersten 22 Studienwochen. Obwohl für 90 % der Studienpopulation mindestens eine Begleitmedikation zur Behandlung des Pruritus dokumentiert ist, wird nicht von verzerrenden Aspekten auf die Wirksamkeitsendpunkte zum Pruritus ausgegangen. Zum einen, weil die Dosierung in den ersten 22 Studienwochen stabil gehalten werden musste, zum anderen, weil ein bestehender mindestens moderater Pruritus ein Einschlusskriterium der Studie war.

Basierend auf dem Studiendesign der Studie ICONIC liegen für die Population der Maralixibat-naiven Personen nur deskriptive Ergebnisse zu Wirksamkeit und Sicherheit von Maralixibat aus den unkontrollierten Titrations- bzw. Erhaltungsperioden vor. Der durchgeführte randomisiert kontrollierte Vergleich in der RW-Phase berücksichtigt nur vorbehandelte Personen, die sich zu Woche 18 noch in der Studie befanden. Da nur 2 Personen die Studie bis zu diesem Zeitpunkt abbrechen, wird die Abweichung im Vergleich zur ITT-Population als geringfügig erachtet. Allerdings ist zu berücksichtigen, dass der randomisierte Vergleich nur die Frage des Effekts einer Weiterbehandlung mit Maralixibat im Vergleich zum Absetzen der Therapie auf ein Placebo nach 18 Behandlungswochen adressiert. Damit ist der Vergleich nicht geeignet, direkte Aussagen zu Therapieeffekten einer neu initiierten Maralixibat-Therapie bei Personen mit ALGS abzuleiten.

Das Verzerrungspotential der 4-wöchigen RW-Phase wird als hoch eingeschätzt. Die Aussagekraft der Ergebnisse ist zusätzlich durch die kurze Dauer des randomisierten Vergleichs und der

geringen Patientenzahl erheblich eingeschränkt. Für die Ergebnisse der nicht-vergleichenden Studienphasen ist von einem hohen Verzerrungspotential auszugehen.

### **Studie MRX-801**

Bei der ergänzend im EPAR berücksichtigten Studie MRX-801 handelt es sich um eine offene, multizentrische Phase-II-Studie zur Untersuchung der Sicherheit und Verträglichkeit von Maralixibat bei der Behandlung von Säuglingen (< 12 Monate) mit cholestatischen Lebererkrankungen (PFIC oder ALGS). Für die Nutzenbewertung sind ausschließlich Daten zum Krankheitsbild ALGS relevant. Nach einer 4-wöchigen Screening-Phase wurde Maralixibat in 6 Wochen auf eine Dosis von maximal 400 µg/kg/Tag oder die höchste tolerierte Dosis auftitriert, die in der anschließenden Erhaltungsphase von Woche 7 bis Woche 13 weiter verabreicht wurde. Eine Weiterbehandlung in der Langzeitverlängerung war mindestens bis zu einem Alter von 12 Monaten vorgesehen. Nach Erreichen dieses Alters erfolgt eine Weiterbehandlung bis zum Eintreten definierter Ereignisse. Primäres Zielkriterium der Studie war die Untersuchung der Sicherheit und Verträglichkeit. Weiterhin waren neben s-BA-Spiegel und Leberwerten auch der ItchRO, die klinische Kratzskala und das Wachstum als Endpunkte definiert. Abweichend von der Studie ICONIC war das Vorhandensein eines cholestatischen Pruritus kein explizites Einschlusskriterium.

Bis zum Interim-Datenschnitt vom 04.05.2022 wurden 8 Säuglinge eingeschlossen und mit Maralixibat behandelt. Alle haben bereits die 13-wöchige Kernstudienphase beendet, wobei 1 Säugling die Studie aufgrund des Entzugs der Einverständniserklärung abbrach. 6 entschieden sich für eine Weiterbehandlung in der langfristigen Behandlungsphase ab Woche 13, von denen sich zum Interim-Datenschnitt noch 5 unter Behandlung befanden. Ein Kleinkind hat die Studie durch den Wechsel in ein „Expanded Access Program“ abgeschlossen. Die mediane Behandlungsdauer liegt bei 137 Tagen. Zur Charakterisierung der Studienpopulation liegen nur wenige Angaben vor. Demnach betrug das mediane Alter zu Baseline 7 Monate und 7 der 8 Säuglinge waren männlichen Geschlechts. Bei den Säuglingen zeigten sich bereits deutlich negative Abweichungen bei den Parametern der körperlichen Entwicklung im Vergleich zur Norm. Informationen hinsichtlich ALGS-Diagnosekriterien und Herkunft der Kinder liegen nicht vor. Auf Basis der klinischen Kratzskala ist bekannt, dass 7 der 8 Studienteilnehmenden zu Baseline unter Pruritus gemäß klinischer Bewertung litten.

Neben Ergebnissen zu UE sind für die Wirksamkeit zum Interim-Datenschnitt lediglich Ergebnisse zur klinischen Kratzskala berichtet. Es ist nicht beschrieben, weshalb keine Angaben zum ItchRO oder der körperlichen Entwicklung vorliegen, da die 13-wöchige Erhaltungsphase der Studie zu diesem Zeitpunkt bereits von allen Studienteilnehmenden abgeschlossen wurde. Abgesehen davon erscheint es allgemein fraglich, wie valide die ALGS-Symptomatik gerade in den ersten Lebensmonaten erfasst werden kann bzw. sich Juckreiz bereits im Verhalten der Kinder bemerkbar macht (Kratzen, inklusive offener Wunden oder Narben). Es ist daher davon auszugehen, dass die Wirksamkeit von Maralixibat bei Säuglingen abschließend nicht bewertet werden kann.

Da es sich bei der Studie MRX-801 um eine einarmige Studie handelt, wird allgemein von einem hohen Verzerrungspotential ausgegangen.

### **Indirekter Vergleich (Studie GALA-MRX-ALGS)**

Zusätzlich zur pivotalen Studie ICONIC legt der pU im Nutzendossier ergänzend die Studie GALA-MRX-ALGS vor, eine Untersuchung zur Langzeitbehandlung mit Maralixibat bei Personen mit ALGS im Vergleich zu einer externen Maralixibat-naiven Kontrollkohorte aus dem GALA-Register, was einem indirekten Vergleich ohne Brückenkompator entspricht. Wie in Kapitel 2.6 ausgeführt, werden die Ergebnisse des indirekten Vergleichs nicht berücksichtigt, da die Patientinnen und Patienten der Interventionsgruppe mehrheitlich aus klinischen Studien rekrutiert wurden, in denen Maralixibat in einer nicht fachinformationskonformen Dosierung von maximal 280 µg/kg/Tag eingesetzt wurde. Separate Auswertungen für die fachinformationskonform behandelte Population liegen nicht vor. Darüber hinaus ist der indirekte Vergleich auch aufgrund einer fehlenden Begründung für die Auswahl der Kontrollkohorte, mangelhafter Angaben zur Vergleichbarkeit der Patientencharakteristika (z. B. der Krankheitsschwere), nicht-verfügbaren Angaben zur Pruritus-Prävalenz in der Kontrollkohorte sowie einer hohen Anzahl fehlender Werte stark limitiert und kann nicht berücksichtigt werden.

### **4.3 Mortalität**

Todesfälle wurden in beiden Studien im Rahmen der Sicherheitserhebung erfasst. Bis zum finalen Datenschnitt waren keine Todesfälle zu verzeichnen.

Eine Interpretation und Bewertung der Daten, insbesondere auch der vergleichenden Studiendaten, zur Mortalität ist aufgrund der kurzen Placebo-kontrollierten Studienphase in der Studie ICONIC nicht möglich und der Effekt von Maralixibat auf die Mortalität kann auf Grundlage der vorgelegten Daten nicht beurteilt werden.

### **4.4 Morbidität**

Für die Endpunktkategorie „Morbidität“ legt der pU Daten zum Pruritus – mittels Tagebuch (ItchRO), Eindruck der Veränderung (PIC/CIC) und klinischer Kratzskala – vor sowie Ergebnisse zu Fatigue, der körperlichen Entwicklung und Xanthomen. Ergänzend wurden Ergebnisse zu Leberfunktion und sBA-Spiegel berichtet.

Die detaillierte Bewertung der Endpunkte ist Kapitel 2.3 zu entnehmen. Zusammenfassend wurden für die Nutzenbewertung Ergebnisse zum Pruritus, mittels ItchRO und der klinischen Kratzskala, sowie der Fatigue und der körperlichen Entwicklung berücksichtigt. Der primäre Endpunkt der Studie ICONIC „Veränderung des Nüchtern-sBA-Spiegels zu Woche 22 gegenüber Woche 18 bei Personen, die zuvor auf die Maralixibat-Behandlung angesprochen haben“ wird aufgrund unklarer Patientenrelevanz des sBA (vgl. Kapitel 2.3.2) lediglich ergänzend im Anhang dargestellt.

Der Endpunkt „Eindruck der Veränderung (PIC/CIC)“ wird in der Nutzenbewertung zur Vermeidung einer Doppelerfassung nicht berücksichtigt, da Pruritus bereits kontinuierlich mittels Tagebuch (ItchRO) erfasst wurde. Zudem liegen für die diesen Endpunkt keine Auswertungen für die RW-Phase vor.

Xanthome können, abhängig vom Grad der Betroffenheit und der Beeinträchtigung, patientenrelevant sein. In der Studie ICONIC sind die Patientinnen und Patienten beeinträchtigende Xanthome zu Baseline bei lediglich 16 % prävalent, sodass die Relevanz einer Verringerung der Xanthome unklar ist. In Verbindung mit der unklaren Validität der Erhebung wird der Endpunkt „Xanthome“ nicht berücksichtigt. Die Leberwerte werden als nicht patientenrelevant erachtet und nicht herangezogen.

## ItchRO

Zur Auswertung des elektronischen Tagebuchs in der betreuerberichteten (ItchRO(Obs)) bzw. patientenberichteten (ItchRO(Pt)) Version liegen eine Vielzahl von Operationalisierungen vor. Primär werden die Responderanalysen in der RW-Phase als relevant erachtet.

Es liegen keine Ergebnisse auf Patientenebene für das Erreichen von Werten im ItchRO  $\leq 1$  Punkt vor. Diese werden jedoch als relevant eingestuft, da damit ein Verschwinden des Pruritus oder eine nur noch leichte Pruritus-Symptomatik verbunden ist. Vom pU wurde lediglich der Anteil der Tage mit einem Wert für die Erhebungen am Morgen (entspricht dem nächtlichen Pruritus) von  $\leq 1$  Punkt im ItchRO deskriptiv und kumuliert für alle Personen ausgewertet. Dabei ist der Anteil in der RW-Phase mit 44 % unter Maralixibat höher als im Placebo-Arm mit 28 %. Auch im ItchRO(Pt) zeigen sich im Interventionsarm anteilig mehr Tage mit Werten  $\leq 1$  Punkt (58 %) verglichen mit Placebo (42 %). In der Screening-Periode sind, zum Vergleich, für jeweils nur 6 % der Tage Erhebungen von  $\leq 1$  Punkt im ItchRO am Morgen dokumentiert. Eine Testung auf statistische Signifikanz erfolgte nicht.

Bei den kontinuierlichen Auswertungen in der RW-Phase (Woche 18 zu Woche 22) zeigt sich hinsichtlich der Veränderung des Pruritus im ItchRO(Obs) anhand des wöchentlichen Durchschnitts bei der Erhebung am Morgen ein statistisch signifikanter Unterschied zugunsten von Maralixibat (LS-Mittelwertdifferenz: -1,48 (95%-KI: [-2,12; -0,84];  $p < 0,0001$ )). Bei der patientenberichteten Version, dem ItchRO(Pt), zeigt sich in dieser Operationalisierung ebenfalls ein statistisch signifikanter Unterschied zugunsten vom Maralixibat (LS-Mittelwertdifferenz: -1,99 (95%-KI: [-3,01; -0,97];  $p = 0,0013$ )). Die klinische Relevanz des Effekts ist jeweils unklar. Konsistente, statistisch signifikante Effekte zeigen sich dabei sowohl bei den weiteren kontinuierlichen Auswertungen (Erhebung am Abend, schwerster Tageswert, gemittelter Tageswert) als auch bei den durchgeführten Sensitivitätsanalysen, wobei neben der Stratifizierungsvariable „sBA-Responder“ auch für weitere Variablen zu Baseline kontrolliert sowie ein MMRM gerechnet wurde. Dies gilt sowohl für den ItchRO(Obs) als auch für den ItchRO(Pt), soweit verfügbar.

Bei den unkontrollierten Daten über den gesamten Studienverlauf zeigt sich in der deskriptiven Darstellung sowohl im ItchRO(Obs) als auch im ItchRO(Pt) in den unterschiedlichen einarmigen Behandlungsphasen ein stetig steigender Anteil an Tagen mit einem Wert von  $\leq 1$  Punkt (geringer oder kein Pruritus) in der Erhebung am Morgen von 6 % zum Screening bis 66 % in der betreuerberichteten Version und 85 % in der patientenberichteten Version für den Zeitraum  $>$  Woche 22 bis Woche 48. Auch hier ist wieder nur der prozentuale Anteil aller Erhebungen der gesamten Studienpopulation berichtet und es ist unklar, wie sich die Tage mit einem Wert von  $\leq 1$  Punkt auf die einzelnen Patientinnen und Patienten verteilen.

Zusätzlich dazu konnte bei den weiteren durchgeführten Responderauswertungen gezeigt werden, dass sowohl zu Woche 18 als auch zu Woche 48 68 % der Studienpopulation eine Reduktion des wöchentlichen Durchschnittswerts der Erhebung am Morgen im ItchRO(Obs) um mindestens 1 Punkt erreicht hatten. In den kontinuierlichen Auswertungen der unkontrollierten Daten über den gesamten Studienverlauf besteht bei einer Skalenspannweite von 0 bis 4 Punkten sowohl im ItchRO(Obs) mit -1,62 Punkten als auch im ItchRO(Pt) mit -2,24 Punkten eine deutliche numerische Reduktionen im wöchentlichen Durchschnittswert für die Erhebung am Morgen. Diese Ergebnisse sind aufgrund fehlender Kontrollgruppe nur eingeschränkt aussagekräftig.

Zusammenfassend zeigen sich während der RW-Phase in den kontinuierlichen Auswertungen des ItchRO konsistent statistisch signifikante Unterschiede zugunsten von Maralixibat, die in allen Operationalisierungen und Sensitivitätsanalysen auf einem Anstieg der Pruritus-Werte in der Placebo-Gruppe basieren. In der Interventionsgruppe blieb die zu Woche 18 eingetretene Reduktion im Pruritus-Schweregrad im Vergleich zu Baseline dagegen weitgehend erhalten. Die

klinische Relevanz dieser Gruppenunterschiede ist jedoch unklar, weshalb ein Vorteil von Maralixibat allein aus den kontinuierlichen Auswertungen nicht abgeleitet werden kann. Die Ergebnisdarstellung der Responder erfolgt rein deskriptiv. Dennoch sind sowohl im Anteil der Tage mit einem Wert  $\leq 1$  Punkt (kein oder geringer Pruritus) als auch dem Anteil der Responder mit einer Reduktion um mindestens 1 Punkt (wöchentliche Durchschnitte) Verbesserungen unter Maralixibat im Vergleich zur Placebo-Gruppe bzw. im Vergleich zum Behandlungsbeginn evident. Gestützt werden diese Ergebnisse auch durch die längeren, einarmigen, unkontrollierten Auswertungen im Vergleich zu Baseline. Dem stehen Unklarheiten hinsichtlich der Validierung des Erhebungsinstruments ItchRO (siehe Kapitel 2.3.2), mögliche Carry-over-Effekte in der Placebo-Gruppe, keine Angaben zu fehlenden Werten sowie das Studiendesign allgemein als Limitationen gegenüber. Für den ItchRO(Pt) ist zudem zu berücksichtigen, dass sich die Ergebnisse nur auf die Subpopulation der  $\geq 5$ -Jährigen beziehen. Da eine Stratifizierung der Randomisierung nach dem Alter nicht vorgesehen war, können aufgrund der kleinen Fallzahl Imbalancen entstehen und folglich ist die Vergleichbarkeit der Behandlungsgruppen unklar. So ist die Population der  $\geq 5$ -Jährigen im Placebo-Arm ( $N = 11$ ) mehr als doppelt so groß wie die der Maralixibat-Gruppe ( $N = 5$ ). Das Verzerrungspotential des ItchRO wird allgemein mit hoch bewertet. In der Gesamtschau ist dennoch von einem Vorteil von Maralixibat gegenüber Placebo auszugehen, wobei die Aussagesicherheit der Ergebnisse durch die kurze Dauer der RW-Phase und die kleine Patientenpopulation erheblichen Unsicherheiten unterliegt.

Daten aus dem Pruritus-Tagebuch ItchRO(Obs) liegen für die Studie MRX-801 nicht vor. Das Instrument wurde jedoch erst wenige Monate vor dem Interim-Datenschnitt in das Protokoll und die Erhebungen in einem Amendment implementiert. Es ist unklar, ob für den ItchRO(Obs) überhaupt Baseline-Daten existieren.

### **Körperliche Entwicklung**

Die körperliche Entwicklung wurde anhand der Veränderungen in den anthropometrischen Parametern „Körpergröße“, „Körpergewicht“ und „BMI“ unter Berücksichtigung standardisierter z-Scores evaluiert. Vergleichende Auswertungen für Veränderung in der RW-Phase liegen nicht vor. Die präsentierten Ergebnisse getrennt für die Behandlungsarme können zur Bewertung des Therapieeffekts von Maralixibat gegenüber Placebo nicht berücksichtigt werden, da sie sich auf die Veränderung von Woche 22 gegenüber Baseline beziehen.

Auf Gesamtstudienebene zeigen sich für die Parameter „Körpergröße“ und „Körpergewicht“ zu Woche 48 im Vergleich zu Baseline leichte numerische Verbesserungen in den z-Scores um 0,18 und 0,02 Punkte. Nach Woche 48 liegen nur noch Daten von weniger als 70 % der Studienteilnehmenden vor, weshalb diese Auswertungen nicht für die Nutzenbewertung berücksichtigt werden.

Aufgrund der unkontrollierten Auswertungen wird das Verzerrungspotential des Endpunkts als hoch bewertet. Zudem ist nicht beschrieben, dass die Erhebung der anthropometrischen Parameter standardisiert erfolgt ist. Die unkontrollierten Daten bis Woche 48 werden als nicht ausreichend erachtet, um die Effekte von Maralixibat auf die körperliche Entwicklung erfassen und bewerten zu können.

Für die Studie MRX-801 liegen keine Ergebnisse zur körperlichen Entwicklung vor.

### **PedsQL-Fatigue**

Mit der PedsQL-Fatigue wurde in der Studie ICONIC ein Zusatzmodul des PedsQL zur Erfassung der Fatigue und Fatigue-assoziiierter Belastungen eingesetzt. Elternberichtete Versionen des Instruments existieren ab einem Alter von 2 Jahren. Bei Kindern und Jugendlichen ab einschließlich 5 Jahren konnte das Instrument zusätzlich in Selbsteinschätzung erhoben werden (PedsQL Child), gegebenenfalls unter Hilfestellung der betreuenden Person. Die Bezugsperiode der eingesetzten Versionen waren jeweils die vergangenen 4 Wochen.

Ausgewertet wurde die PedsQL-Fatigue anhand von Mittelwertdifferenzen im Vergleich zu Baseline bzw. zu Woche 18 für die RW-Phase. Responderanalysen unter Berücksichtigung einer Responseschwelle von 15 % der Skalenspannweite liegen nicht vor. Die klinische Relevanz der gezeigten Effekte ist damit unklar.

In der RW-Phase zeigten sich jeweils keine statistisch signifikanten Unterschiede zwischen den Behandlungsgruppen in der Eltern- und der Kinderversion. Zwar ist insbesondere in der elternberichteten Version der PedsQL-Fatigue eine deutliche numerische Verbesserung um 20 Punkte zu Woche 18 und Woche 48 im Vergleich zu Baseline berichtet. Das Absetzen von Maralixibat während der RW-Phase führt jedoch nicht zu einer statistisch signifikanten Verschlechterung im Placebo-Arm im Vergleich zur Weiterbehandlung im Maralixibat-Arm.

Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass der Bezugszeitraum der eingesetzten PedsQL-Version mit 4 Wochen der gesamten Dauer der RW-Phase entspricht. Gerade in den ersten Tagen nach dem Absetzen von Maralixibat zu Woche 18 ist in der Placebo-Gruppe von Carry-over-Effekten auszugehen. Sofern von einer graduellen Verschlechterung der Fatigue im Placebo-Arm auszugehen ist (analog zum ItchRO), können die Effekte eines Absetzens von Maralixibat im Vergleich zu einer Weiterbehandlung mit der vorliegenden Operationalisierung der Fatigue möglicherweise nicht vollständig erfasst werden.

Bei beiden Versionen des Instruments ist darüber hinaus einschränkend zu berücksichtigen, dass sich die Ergebnisse nur auf die Subpopulation der  $\geq 2$ -Jährigen (Elternversion) bzw. der  $\geq 5$ -Jährigen (Kinderversion) der Studie ICONIC beziehen. Da eine Stratifizierung der Randomisierung nach dem Alter nicht vorgesehen war, können aufgrund der kleinen Fallzahl Imbalancen entstehen und die Vergleichbarkeit der Behandlungsgruppen ist unklar. So ist die Population der  $\geq 5$ -Jährigen im Placebo-Arm ( $N = 11$ ) mehr als doppelt so groß wie die der Maralixibat-Gruppe ( $N = 5$ ). Bei den  $\geq 2$ -Jährigen ist der Unterschied mit 10 Personen im Maralixibat-Arm und 14 in der Placebo-Gruppe nicht ganz so ausgeprägt.

Insgesamt weisen die Ergebnisse der Fatigue anhand des PedsQL ein hohes Verzerrungspotential während der RW-Phase auf. Die nicht-vergleichenden Auswertungen über den gesamten Studienverlauf werden ebenfalls als potentiell hochverzerrt erachtet. Weiterhin ist aufgrund der relativ kurzen vergleichenden Studienphase von nur 4 Wochen die Aussagesicherheit der Ergebnisse erheblich eingeschränkt und die Effekte von Maralixibat auf die Fatigue können nicht abschließend bewertet werden.

## 4.5 Lebensqualität

Lebensqualität wurde in der Studie ICONIC mittels PedsQL erhoben, in der jeweiligen altersgerechten Version. Die elternberichtete Version sollte für alle Studienteilnehmenden beantwortet werden. Bei Kindern und Jugendlichen ab einschließlich 5 Jahren konnte das Instrument analog zur PedsQL-Fatigue zusätzlich in Selbsteinschätzung erhoben werden (PedsQL Child), gegebenenfalls unter Hilfestellung der betreuenden Person. Bezugsperiode der eingesetzten Versionen waren jeweils die vergangenen 4 Wochen.

Auch der PedsQL wurde anhand von Mittelwertdifferenzen im Vergleich zu Baseline bzw. zu Woche 18 für die RW-Phase ausgewertet, sodass die klinische Relevanz der gezeigten Effekte unklar ist.

Bei etwas höheren und damit besseren Werten zu Baseline und Woche 18 in der Maralixibat-Gruppe im Vergleich zu Placebo im PedsQL (Parent) zeigten sich bei der Veränderung in der RW-Phase (Woche 18 zu Woche 22) in beiden Behandlungsgruppen leichte Verschlechterungen in der Lebensqualität. Der Unterschied war zwischen den Behandlungsarmen statistisch nicht signifikant und auch im PedsQL (Child) zeigten sich keine statistisch signifikanten Unterschiede. Zwar liegen für beide Versionen numerische Verbesserungen zu Woche 18 und Woche 48 im Vergleich zu Baseline vor, das Absetzen von Maralixibat während der RW-Phase führt jedoch nicht zu einer statistisch signifikanten Verschlechterung im Placebo-Arm gegenüber dem Maralixibat-Arm.

Die im Abschnitt zur PedsQL-Fatigue thematisierten Limitationen bezüglich möglicher Carry-over-Effekte treffen auch auf den PedsQL zu. Auch die Problematik möglicher Imbalancen zwischen den Behandlungsgruppen aufgrund der betrachteten Subpopulation der  $\geq 5$ -Jährigen besteht analog zur PedsQL-Fatigue.

Insgesamt weisen die Ergebnisse der Lebensqualität mittels PedsQL ein hohes Verzerrungspotential während der RW-Phase auf. Die nicht vergleichenden Auswertungen über den gesamten Studienverlauf werden ebenfalls als potentiell hochverzerrt erachtet. Weiterhin ist aufgrund der relativ kurzen vergleichenden Studienphase von nur 4 Wochen die Aussagesicherheit der Ergebnisse erheblich eingeschränkt und die Effekte von Maralixibat auf die Lebensqualität können nicht abschließend bewertet werden.

Lebensqualität wurde in der Studie MRX-801 nicht erhoben.

## 4.6 Sicherheit

In Bezug auf die UE bleibt unklar, ob auf Basis der Bestimmungen im Protokoll ausreichend zwischen Grunderkrankung und Nebenwirkungen unterschieden werden kann. Unter Berücksichtigung der vorliegenden Sicherheitsdaten in der RW-Phase ist jedoch anzunehmen, dass auch Symptomatik als UE gewertet wurde. So sind in der RW-Phase bei 5 Personen (31 %) aus dem Placebo-Arm ein UE des PT „Pruritus“ berichtet verglichen mit nur 1 Person (6 %) aus dem Interventionsarm. Eine Erfassung erneut aufflammender Symptomatik nach dem Absetzen von Maralixibat erscheint plausibel. Bei der Interpretation der UE einzelner Studienphasen (z. B. RW-Phase oder Woche > 22 bis Woche 48) ist darüber hinaus zu beachten, dass nur in der jeweiligen Studienphase neu aufgetretene oder sich verschlimmernde Ereignisse als UE gewertet wurden. Ein mögliches Abklingen oder eine Verringerung von UE, wie sie beim Absetzen von Maralixibat zu erwarten sind, wird in der Dokumentation und Auswertung der UE nicht erfasst.

Aufgrund der genannten Limitationen wird die vergleichende Darstellung der UE in der RW-Phase als nicht valide erachtet und in der vorliegenden Nutzenbewertung nicht berücksichtigt.

Dennoch ist auch für die einarmige, unkontrollierte Darstellung der Inzidenz der UE davon auszugehen, dass Symptomatik als UE gewertet wurde.

In Modul 4 präsentiert der pU Auswertung aller UE in den unterschiedlichen Studienphasen, wobei UE über einer Dosis von 400 µg/kg/Tag hier nicht berücksichtigt werden. Dabei bezieht sich der Anteil der Personen mit Ereignis auf diejenigen Personen, die sich jeweils noch in der Studie befanden, z. B. N = 23 in der offenen Behandlungsverlängerung. Zudem werden lediglich alle UE gewertet, die in der jeweiligen Studienphase neu auftreten, sodass unklar bleibt wie viele Personen insgesamt unter Maralixibat ein UE bzw. ein UE bestimmter SOC oder PT erleiden. Dieser Darstellung wird nicht gefolgt. Für die Nutzenbewertung sind prinzipiell alle UE relevant, die bis einschließlich einer Dosis von 400 µg/kg/Tag auftraten, unabhängig von der Studienphase. Solche Auswertungen liegen für die Studie ICONIC nicht vor, weshalb bei den Gesamtraten alle UE bis zu einer Dosis von < 400 µg/kg/Tag sowie bei einer Dosis von 400 µg/kg/Tag getrennt dargestellt wurden. Unter Dosierungen von < 400 µg/kg/Tag erlitten 87 % der Personen mindestens ein UE. In einem Fall war das UE schwer und eine weitere Person brach die Studie aufgrund von UE ab. Unter Maralixibat 400 µg/kg/Tag ereignete sich bei fast allen Personen mindestens 1 UE (97 %). Jeweils 9 Personen erlitten jeweils ein schweres UE und ein SUE, wobei insgesamt 5 Studienabbrüche aufgrund von UE dokumentiert sind.

Für UE nach SOC und PT, die sich bei Dosierungen < 400 µg/kg/Tag ereignen, liegen keine Auswertungen vor. Die häufigsten UE unter Maralixibat 400 µg/kg/Tag waren die PT „Abdominalschmerz“, „Diarrhö“ und „Erbrechen“ aus der SOC „Erkrankungen des Gastrointestinaltrakts“ (74 %), von denen jeweils über 36 % der Studienpopulation betroffen war. Auch das PT „Pyrexie“ aus der SOC „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ war mit knapp 50 % häufig zu verzeichnen. Bei den schweren UE sticht kein SOC oder PT als besonders häufig heraus. Ausschließlich die PT „Abdominalschmerz“ und „Alanin-Aminotransferase erhöht“ (jeweils 7 %) ereigneten sich bei mehr als 5 % der Sicherheitspopulation. SUE treten hauptsächlich in den SOC „Erkrankungen des Gastrointestinaltrakts“ (10 %) und „Infektionen und parasitäre Erkrankungen“ (16 %) auf. SUE eines PT treten in keinem Fall bei mehr als 5 % der Sicherheitspopulation auf. Bei den UE von besonderem Interesse waren insbesondere „Diarrhö“ (48 %) und „Ereignisse im Zusammenhang mit Mängeln der fettlöslichen Vitamine“ (20 %) häufiger zu verzeichnen. Ergebnisse für schwere und schwerwiegende UE von besonderem Interesse liegen nicht vor. Es wird darauf hingewiesen, dass die Definition der UE von besonderem Interesse nach Entblindung der Studie im SAP geändert worden ist. Eine Rationale hierfür ist nicht angegeben.

In der Studie MRX-801 ereignete sich bei 88 % der Patientenpopulation mindestens 1 UE. Schwere UE und SUE traten jeweils bei 4 Kleinkindern (50 %) auf, wobei die schweren UE und SUE nach SOC und PT vergleichbar sind und mit den PT „Infantile Koliken“, „Pyrexie“, „Weinen“, „Corona-Infektion“, „Varizella“ und „Virusinfektion“ bei jeweils 1 Person insbesondere die SOC „Infektionen und parasitäre Erkrankungen“ sowie „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ betrafen. UE von besonderem Interesse ereigneten sich bei keinem Kleinkind.

Vor dem Hintergrund eines fehlenden validen Vergleichs, des geringen Stichprobenumfangs und der Einschränkungen bei der Erhebung und Auswertung der Sicherheitsdaten hinsichtlich der Erfassung von Krankheitssymptomatik, ist eine abschließende Bewertung zur Sicherheit von Maralixibat bei Kindern und Jugendlichen mit ALGS nicht möglich.

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

Die Angaben der Fachinformation sind zu berücksichtigen.

## 6 Zusammenfassung der Nutzenbewertung

Maralixibat (Livmarli®) ist zugelassen zur Behandlung des cholestatischen Pruritus bei Patientinnen und Patienten mit Alagille-Syndrom (ALGS) ab dem Alter von 2 Monaten. Die Nutzenbewertung von Maralixibat basiert auf der zulassungsbegründenden Studie ICONIC, einer offenen Langzeitstudie mit einer doppelblinden, Placebo-kontrollierten, randomisierten Absetzphase bei Kindern mit ALGS, zur Untersuchung der Sicherheit und Wirksamkeit von Maralixibat. Zusätzlich wurden Interim-Ergebnisse der Studie MRX-801 berücksichtigt, einer offenen, multi-zentrischen Phase-II-Studie zur Untersuchung der Sicherheit und Verträglichkeit von Maralixibat bei der Behandlung von Säuglingen (< 12 Monate) mit cholestatischen Lebererkrankungen (PFIC oder ALGS).

Die Ergebnisse der in der Nutzenbewertung berücksichtigten Endpunkte der Studien ICONIC und MRX-801 werden in den folgenden Tabellen unter Berücksichtigung des Verzerrungspotentials zusammengefasst. Es wird dargestellt, ob unter Maralixibat ein statistisch signifikant positiver Effekt bei hohem oder unklarem (↑) oder niedrigem (↑↑) Verzerrungspotential bzw. ein statistisch signifikant negativer Effekt bei hohem oder unklarem (↓) oder niedrigem (↓↓) Verzerrungspotential bzw. kein Unterschied (↔) gezeigt werden konnte.

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

Tabelle 34: Zusammenfassende Darstellung der vergleichenden Ergebnisse der Studie ICONIC – RW-Phase

| Studie ICONIC <sup>1)</sup><br>Darstellung der Ergebnisse   | Maralixibat<br>N = 13 <sup>2)</sup> |                   |  |              | Placebo<br>N = 16 <sup>2)</sup> |                   |  |             | Intervention vs. Kontrolle                               | Effekt |
|---|-------------------------------------|-------------------|--|--------------|---------------------------------|-------------------|--|-------------|--|--------|
| <b>Mortalität</b>   |                                     |                   |  |              |                                 |                   |  |             |  |        |
|   | N <sup>3)</sup>                     | Ereignisse, n (%) |  |              | N <sup>3)</sup>                 | Ereignisse, n (%) |  |             |  |        |
| Todesfälle  | 13                                  | 0                 |  |              | 16                              | 0                 |  |             | -  | n. b.  |
| <b>Morbidität</b>   |                                     |                   |  |              |                                 |                   |  |             |  |        |
| <b>ItchRO(Obs)<sup>7)</sup></b>                             | N <sup>4)</sup>                     | Ereignisse, n (%) |  |              | N <sup>4)</sup>                 | Ereignisse, n (%) |  |             | Effektschätzer   |        |
| Anteil Tage mit einem Wert<br>≤ 1 Punkt; Erhebung am Morgen | 435                                 | 190 (44)          |  |              | 455                             | 127 (28)          |  |             | k. A.  | n. b.  |
| Anteil Personen mit einem<br>Wert ≤ 1 Punkt                 | k. A.                               | k. A.             |  |              | k. A.                           | k. A.             |  |             | k. A.  | n. b.  |
| <b>ItchRO(Pt)<sup>7)</sup></b>                              | N <sup>5)</sup>                     | Ereignisse, n (%) |  |              | N <sup>5)</sup>                 | Ereignisse, n (%) |  |             | Effektschätzer   |        |
| Anteil Tage mit einem Wert<br>≤ 1 Punkt; Erhebung am Morgen | 233                                 | 136 (58)          |  |              | 258                             | 109 (42)          |  |             | k. A.  | n. b.  |
| Anteil Personen mit einem<br>Wert ≤ 1 Punkt                 | k. A.                               | k. A.             |  |              | k. A.                           | k. A.             |  |             | k. A.  | n. b.  |
| <b>ItchRO(Obs)<sup>7)</sup></b>                             | Woche 18                            |                   | Veränderung in der<br>RW-Phase <sup>6)</sup> |              | Woche 18                        |                   | Veränderung in der<br>RW-Phase <sup>6)</sup> |             | LS-Mittelwertdifferenz<br>[95%-KI]; p-Wert <sup>6)</sup> |        |
|   | N <sup>3)</sup>                     | MW (SD)           | N <sup>3)</sup>                              | LS MW (SE)   | N <sup>3)</sup>                 | MW (SD)           | N <sup>3)</sup>                              | LS MW (SE)  |  |        |
| Erhebung am Morgen  | 13                                  | 1,30 (0,86)       | 12   | 0,22 (0,23)  | 16                              | 1,13 (0,85)       | 16   | 1,70 (0,20) | -1,48 [-2,12; -0,84]; < 0,0001                           | ↑      |
| <b>ItchRO(Pt)<sup>7)</sup></b>                              | Woche 18                            |                   | Veränderung in der<br>RW-Phase <sup>6)</sup> |              | Woche 18                        |                   | Veränderung in der<br>RW-Phase <sup>6)</sup> |             | LS-Mittelwertdifferenz<br>[95%-KI]; p-Wert <sup>6)</sup> |        |
|   | N <sup>3)</sup>                     | MW (SD)           | N <sup>3)</sup>                              | LS MW (SE)   | N <sup>3)</sup>                 | MW (SD)           | N <sup>3)</sup>                              | LS MW (SE)  |  |        |
|   | N = 5 <sup>8)</sup>                 |                   |  |              | N = 11 <sup>8)</sup>            |                   |  |             |  |        |
| Erhebung am Morgen  | 5                                   | 0,77 (0,76)       | 5  | -0,15 (0,37) | 9                               | 0,87 (0,88)       | 9  | 1,84 (0,28) | -1,99 [-3,01; -0,97]; < 0,0013                           | ↑      |

| Studie ICONIC <sup>1)</sup><br>Darstellung der Ergebnisse | Maralixibat<br>N = 13 <sup>2)</sup> |               |  |              | Placebo<br>N = 16 <sup>2)</sup> |               |  |               | Intervention vs. Kontrolle                               | Effekt |
|---|-------------------------------------|---------------|--|--------------|---------------------------------|---------------|--|---------------|--|--------|
|   | Woche 18                            |               | Veränderung in der<br>RW-Phase <sup>6)</sup> |              | Woche 18                        |               | Veränderung in der<br>RW-Phase <sup>6)</sup> |               |  |        |
| PedsQL-Fatigue <sup>9)</sup>                              | N <sup>3)</sup>                     | MW (SD)       | N <sup>3)</sup>                              | LS MW (SE)   | N <sup>3)</sup>                 | MW (SD)       | N <sup>3)</sup>                              | LS MW (SE)    | LS-Mittelwertdifferenz<br>[95%-KI]; p-Wert <sup>6)</sup> |        |
|   | N = 10 <sup>10)</sup>               |               |  |              | N = 14 <sup>10)</sup>           |               |  |               |  |        |
| Elternversion   | 9                                   | 74,85 (11,01) | 9  | -2,96 (6,05) | 14                              | 71,03 (14,88) | 12   | -16,99 (5,24) | 14,03 [-2,78; 30,84]; 0,0966                             | ↔      |
|   | N = 5 <sup>8)</sup>                 |               |  |              | N = 11 <sup>8)</sup>            |               |  |               |  |        |
| Kinderversion   | 5                                   | 61,94 (13,49) | 5  | 1,05 (7,85)  | 10                              | 61,59 (10,48) | 9  | -8,53 (5,85)  | 9,57 [-11,97; 31,11]; 0,3490                             | ↔      |
| <b>Lebensqualität</b>                                     |                                     |               |  |              |                                 |               |  |               |  |        |
| PedsQL <sup>9)</sup>                                      | Woche 18                            |               | Veränderung in der<br>RW-Phase <sup>6)</sup> |              | Woche 18                        |               | Veränderung in der<br>RW-Phase <sup>6)</sup> |               | LS-Mittelwertdifferenz<br>[95%-KI]; p-Wert <sup>6)</sup> |        |
|   | N <sup>3)</sup>                     | MW (SD)       | N <sup>3)</sup>                              | LS MW (SE)   | N <sup>3)</sup>                 | MW (SD)       | N <sup>3)</sup>                              | LS MW (SE)    |  |        |
| Elternversion   | 12                                  | 74,80 (10,37) | 12   | -6,69 (4,51) | 16                              | 68,24 (16,43) | 16   | -9,03 (3,89)  | 2,33 [-10,08; 14,75]; 0,7018                             | ↔      |
|   | N = 5 <sup>8)</sup>                 |               |  |              | N = 11 <sup>8)</sup>            |               |  |               |  |        |
| Kinderversion   | 5                                   | 66,74 (11,07) | 5  | 4,35 (4,46)  | 10                              | 70,43 (10,10) | 9  | -3,50 (3,31)  | 7,85 [-4,44; 20,13]; 0,1874                              | ↔      |

<sup>1)</sup> Finaler Datenschnitt vom 21.08.2020.

<sup>2)</sup> RW-Population bestehend aus allen randomisierten Personen, die mind. eine Dosis Maralixibat erhalten haben.

<sup>3)</sup> Anzahl berücksichtigter Personen.

<sup>4)</sup> Kumulierte Anzahl der Erhebungen in der RW-Periode bei allen Personen aus der RW-Population.

<sup>5)</sup> Kumulierte Anzahl der Erhebungen in der RW-Periode bei allen Personen aus der RW-Population, die den ItchRO(Pt) beantworteten.

<sup>6)</sup> Gemischtes ANCOVA-Modell mit „Behandlungsgruppe“ als festen Effekt und „Baseline-Wert“ als Kovariate.

<sup>7)</sup> Skala von 0 bis 4. Eine höhere Punktzahl repräsentiert einen stärkeren Pruritus.

<sup>8)</sup> Subgruppe der RW-Population im Alter von  $\geq 5$  Jahren.

<sup>9)</sup> Skala von 0 bis 100: Eine höhere Punktzahl repräsentiert eine bessere Lebensqualität bzw. Fatigue.

<sup>10)</sup> Subgruppe der RW-Population im Alter von  $\geq 2$  Jahren.

Abkürzungen: ANCOVA: Analysis of Covariance; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; n. b.: nicht beurteilbar; PedsQL: Pediatric Quality of Life Inventory; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; RW: Randomisierte Absetzphase; SD: Standardabweichung; SE: Standardfehler.

*Tabelle 35: Zusammenfassende Darstellung der Ergebnisse der Studie ICONIC – Unkontrollierte Maralixibat-Behandlung*

| <b>Studie ICONIC<sup>1)</sup><br/>Darstellung der Ergebnisse</b>                   | <b>Maralixibat<br/>N = 31</b> |                                     |
|--|-------------------------------|-------------------------------------|
| <b>Mortalität</b>  |                               |                                     |
|  | <i>N</i>                      | <i>Personen mit Ereignis, n (%)</i> |
| Todesfälle   | 31                            | 0                                   |
| <b>Morbidität</b>  |                               |                                     |
| <b>ItchRO</b>  |                               |                                     |
| ItchRO(Obs): Wöchentlicher Durchschnitt der Erhebung am Morgen – Anteil ≤ 1 Punkt  | <i>N</i>                      | <i>Personen mit Ereignis, n (%)</i> |
| Zu Woche 48  | 31                            | k. A.                               |
| ItchRO(Pt): Wöchentlicher Durchschnitt der Erhebung am Morgen – Anteil ≤ 1 Punkt   | <i>N</i>                      | <i>Personen mit Ereignis, n (%)</i> |
| Zu Woche 48  | 14                            | k. A.                               |
| ItchRO(Obs): Wöchentlicher Durchschnitt der Erhebung am Morgen                     | <i>N</i>                      | <i>MW (SD)</i>                      |
| Woche 48<br>Absoluter Wert<br>Absolute Veränderung                                 | 28                            | 1,28 (1,14)<br>-1,62 (1,30)         |
| ItchRO(Pt): Wöchentlicher Durchschnitt der Erhebung am Morgen                      | <i>N</i>                      | <i>MW (SD)</i>                      |
| Woche 48<br>Absoluter Wert<br>Absolute Veränderung                                 | 14                            | 0,65 (0,76)<br>-2,25 (1,01)         |
| ItchRO(Obs): Wöchentlicher Durchschnitt der Erhebung am Morgen – Responderanalysen | <i>N</i>                      | <i>Personen mit Ereignis, n (%)</i> |
| Verbesserung ≥ 1 Punkt zu Woche 48   | 31                            | 21 (68)                             |
| ItchRO(Pt): Wöchentlicher Durchschnitt der Erhebung am Morgen – Responderanalysen  | <i>N</i>                      | <i>Personen mit Ereignis, n (%)</i> |
| Verbesserung ≥ 1 Punkt zu Woche 48   | 14                            | k. A.                               |
| <b>Körperliche Entwicklung<sup>2)</sup></b>  | <i>N<sup>3)</sup></i>         | <i>MW (SD)</i>                      |
| <b>Körpergröße (z-Score)</b>   |                               |                                     |
| Woche 48<br>Absoluter Wert<br>Absolute Veränderung                                 | 28                            | -1,436 (1,127)<br>0,178 (0,501)     |
| <b>Körpergewicht (z-Score)</b>   |                               |                                     |
| Woche 48<br>Absoluter Wert<br>Absolute Veränderung                                 | 28                            | -1,517 (0,970)<br>0,018 (0,422)     |

| <b>Studie ICONIC<sup>1)</sup><br/>Darstellung der Ergebnisse</b> | <b>Maralixibat<br/>N = 31</b> |                                     |
|--|-------------------------------|-------------------------------------|
| <b>PedsQL-Fatigue (Parent)</b>                                   | <i>N</i>                      | <i>MW (SD)</i>                      |
| Woche 48   | 21                            | N = 22 <sup>4)</sup>                |
| Absoluter Wert   |                               | 73,99 (15,46)                       |
| Absolute Veränderung   |                               | 20,30 (24,87)                       |
| <b>PedsQL-Fatigue (Child)</b>                                    | <i>N</i>                      | <i>MW (SD)</i>                      |
| Woche 48   | 13                            | N = 14 <sup>5)</sup>                |
| Absoluter Wert   |                               | 64,88 (19,17)                       |
| Absolute Veränderung,  |                               | 5,66 (13,91)                        |
| <b>Lebensqualität</b>  |                               |                                     |
| <b>PedsQL (Parent)</b>   | <i>N</i>                      | <i>MW (SD)</i>                      |
| Woche 48   | 27                            |                                     |
| Absoluter Wert   |                               | 68,34 (15,50)                       |
| Absolute Veränderung   |                               | 8,94 (18,74)                        |
| <b>PedsQL (Child)</b>  | <i>N</i>                      | <i>MW (SD)</i>                      |
| Woche 48   | 14                            | N = 14 <sup>5)</sup>                |
| Absoluter Wert   |                               | 69,88 (17,69)                       |
| Absolute Veränderung   |                               | 4,35 (15,15)                        |
| <b>Sicherheit<sup>6)7)</sup></b>                                 | <i>N</i>                      | <i>Personen mit Ereignis, n (%)</i> |
| Schwere UE <sup>8)</sup>   | 31                            | k. A.                               |
| SUE  | 31                            | 9 (29)                              |
| Therapieabbruch aufgrund von UE                                  | 31                            | 6 (19)                              |

<sup>1)</sup> Finaler Datenschnitt vom 21.08.2020.

<sup>2)</sup> Skala von 0 bis 4: Eine höhere Punktzahl repräsentiert einen stärkeren Pruritus.

<sup>3)</sup> Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden.

<sup>4)</sup> Subgruppe der ITT-Population im Alter von  $\geq 2$  Jahren.

<sup>5)</sup> Subgruppe der ITT-Population im Alter von  $\geq 5$  Jahren.

<sup>6)</sup> Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind in Kapitel 3.5 dargestellt.

<sup>7)</sup> Bei der Auswertung sind alle UE relevant, die unter einer fachinformationskonformen Dosis von  $\leq 400$   $\mu\text{g}/\text{kg}/\text{Tag}$  aufgetreten sind. Diese Angaben sind den Unterlagen des pU nicht zu entnehmen. Für die SUE und die Anzahl der Therapieabbrüche aufgrund von UE können die Ereignisse den Ergebnistabellen des pU entnommen werden.

<sup>8)</sup> CTCAE-Grad  $\geq 3$ .

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; ItchRO: Itch Reported Outcome; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); k. A.: keine Angabe; MW: Mittelwert; PedsQL: Pediatric Quality of Life Inventory; PedsQL-Fatigue: Pediatric Quality of Life Inventory – Multidimensional Fatigue Scale; pU: pharmazeutischer Unternehmer; SD: Standardabweichung; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Tabelle 36: Zusammenfassende Darstellung der Ergebnisse der Studie MRX-801

| Studie MRX-801 <sup>1)</sup><br>Darstellung der Ergebnisse | Maralixibat<br>N = 8 |                              |
|--|----------------------|------------------------------|
| <b>Mortalität</b>  |                      |                              |
|  | N <sup>2)</sup>      | Personen mit Ereignis, n (%) |
| Todesfälle   | 8                    | 0                            |
| <b>Sicherheit<sup>3)</sup></b>                             | N <sup>2)</sup>      | Personen mit Ereignis, n (%) |
| Schwere UE <sup>4)</sup>                                   | 8                    | 4 (50)                       |
| SUE  | 8                    | 4 (50)                       |
| Therapieabbruch aufgrund von UE                            | 8                    | 0                            |

<sup>1)</sup> Datenschnitt vom 04.05.2022.

<sup>2)</sup> Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden.

<sup>3)</sup> Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind in Kapitel 3.5 dargestellt.

<sup>4)</sup> CTCAE-Grad  $\geq 3$ .

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## Referenzen

1. **Arbor Research Collaborative for Health.** Longitudinal study of genetic causes of intra-hepatic cholestasis (LOGIC) [online]. NCT00571272. In: ClinicalTrials.gov. Last update posted: 22.11.2022. [Zugriff: 07.03.2023].  
URL: <https://www.clinicaltrials.gov/ct2/show/NCT00571272>.
2. **Center for Drug Evaluation and Research.** Livmarli (maralixibat): application number 214662Orig1s000; integrated review [online]. Silver Spring (USA): U.S. Food and Drug Administration (FDA); 2021. [Zugriff: 05.04.2023]. URL: [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2021/214662Orig1s000IntegratedR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2021/214662Orig1s000IntegratedR.pdf).
3. **Centers for Disease Control and Prevention (CDC).** Growth chart training: the SAS program for CDC growth charts that includes the extended BMI calculations; SAS program for 2000 CDC growth charts (ages 0-20 years) [online]. Updated 09.01.2023. Atlanta (USA): CDC. [Zugriff: 05.04.2023].  
URL: <https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>.
4. **Emerick KM, Whittington PF.** Partial external biliary diversion for intractable pruritus and xanthomas in Alagille syndrome. *Hepatology* 2002;35(6):1501-1506.
5. **European Medicines Agency (EMA).** Livmarli: European public assessment report EMEA/H/C/005857/0000 [online]. Amsterdam (NED): EMA; 2022. [Zugriff: 06.04.2023].  
URL: [https://www.ema.europa.eu/documents/assessment-report/livmarli-epar-public-assessment-report\\_en.pdf](https://www.ema.europa.eu/documents/assessment-report/livmarli-epar-public-assessment-report_en.pdf).
6. **European Medicines Agency (EMA).** Maralixibat (Livmarli): Zusammenfassung der Merkmale des Arzneimittels (EPAR Product Information) [online]. Amsterdam (NED): European Medicines Agency (EMA); 2022. [Zugriff: 06.04.2023].  
URL: [https://www.ema.europa.eu/en/documents/product-information/livmarli-epar-product-information\\_de.pdf](https://www.ema.europa.eu/en/documents/product-information/livmarli-epar-product-information_de.pdf).
7. **Gonzales E, Hardikar W, Stormon M, Baker A, Hierro L, Gliwicz D, et al.** Efficacy and safety of maralixibat treatment in patients with Alagille syndrome and cholestatic pruritus (ICONIC): a randomised phase 2 study. *Lancet* 2021;398(10311):1581-1592.
8. **Grindler DJ, Blank SJ, Schulz KA, Witsell DL, Lieu JE.** Impact of otitis media severity on children's quality of life. *Otolaryngol Head Neck Surg* 2014;151(2):333-340.
9. **Kamath BM, Abetz-Webb L, Kennedy C, Hepburn B, Gauthier M, Johnson N, et al.** Development of a novel tool to assess the impact of itching in pediatric cholestasis. *Patient* 2018;11(1):69-82.
10. **Kamath BM, Baker A, Houwen R, Todorova L, Kerkar N.** Systematic review: the epidemiology, natural history, and burden of alagille syndrome. *J Pediatr Gastroenterol Nutr* 2018;67(2):148-156.
11. **Kamath BM, Goldstein A, Howard R, Garner W, Vig P, Marden JR, et al.** Maralixibat treatment response in alagille syndrome is associated with improved health-related quality of life. *J Pediatr* 2023;252:68-75.
12. **Kamath BM, Spino C, McLain R, Magee JC, Fredericks EM, Setchell KD, et al.** Unraveling the relationship between itching, scratch scales, and biomarkers in children with alagille syndrome. *Hepatology Communications* 2020;4(7):1012-1018.

13. **Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.** Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-766.
14. **Kohut TJ, Gilbert MA, Loomes KM.** Alagille syndrome: a focused review on clinical features, genetics, and treatment. *Semin Liver Dis* 2021;41(4):525-537.
15. **Lykavieris P, Hadchouel M, Chardot C, Bernard O.** Outcome of liver disease in children with Alagille syndrome: a study of 163 patients. *Gut* 2001;49(3):431-435.
16. **Magee J.** Longitudinal study of genetic causes of intrahepatic cholestasis (LOGIC); version 4, dataset [online]. Bethesda (USA): NIDDK Central Repository; 2023. [Zugriff: 01.03.2023]. URL: <https://doi.org/10.58020/dbbh-rt9>.
17. **Mapi Research Trust.** Scaling and scoring for the acute and standard versions of the pediatric quality of life inventory: PedsQL; version 21.3 [online]. Lyon (FRA): Mapi Research Trust; 2023. [Zugriff: 05.04.2023]. URL: <https://www.pedsq.org/PedsQL-Scoring.pdf>.
18. **McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA, et al.** NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am J Hum Genet* 2006;79(1):169-173.
19. **Mirum Pharmaceuticals.** Open-label, phase 2 study to evaluate the safety and tolerability of maralixibat in the treatment of infants with cholestatic liver diseases including progressive familial intrahepatic cholestasis and Alagille syndrome (MRX-801): interim clinical study report [unveröffentlicht]. 20.07.2022.
20. **Mirum Pharmaceuticals.** Open-label, phase 2 study to evaluate the safety and tolerability of maralixibat in the treatment of infants with cholestatic liver diseases including progressive familial intrahepatic cholestasis and Alagille syndrome (MRX-801): study protocol and protocol amendments [unveröffentlicht]. 2022.
21. **Mirum Pharmaceuticals.** A Study to evaluate the safety and tolerability of maralixibat in infant participants with cholestatic liver diseases including progressive familial intrahepatic cholestasis (PFIC) and alagille syndrome (ALGS); (RISE) [online]. NCT04729751. In: *ClinicalTrials.gov*. Last update posted: 29.03.2023. [Zugriff: 30.03.2023]. URL: <https://clinicaltrials.gov/ct2/show/NCT04729751>.
22. **Mirum Pharmaceuticals International.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Maralixibat (Livmarli) zur Behandlung des cholestaticen Pruritus bei Patienten mit Alagille-Syndrom (ALGS) ab dem Alter von 2 Monaten; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 13.01.2023.
23. **Mirum Pharmaceuticals International.** Long-term outcomes of long-term maralixibat treatment in patients with alagille syndrome compared with an external control cohort from the global alagille alliance (GALA) clinical research database; integrated clinical and statistical report [unveröffentlicht]. 2021.
24. **Mirum Pharmaceuticals International.** Long-term, open-label study with a double-blind, placebo-controlled, randomized drug withdrawal period of LUM001, an apical sodium-dependent bile acid transporter inhibitor (ASBTi), in patients with alagille syndrome: study LUM001-304; clinical study protocol [unveröffentlicht]. 08.02.2019.

25. **Mirum Pharmaceuticals International.** Long-term, open-label study with a double-blind, placebo-controlled, randomized drug withdrawal period of LUM001, an apical sodium-dependent bile acid transporter inhibitor (ASBTi), in patients with alagille syndrome: study LUM001-304; clinical study report [unveröffentlicht]. 11.11.2020.
26. **Mirum Pharmaceuticals International.** Long-term, open-label study with a double-blind, placebo-controlled, randomized drug withdrawal period of LUM001, an apical sodium-dependent bile acid transporter inhibitor (ASBTi), in patients with alagille syndrome: study LUM001-304; interim clinical study report [unveröffentlicht]. 13.10.2020.
27. **Mirum Pharmaceuticals International.** Long-term, open-label study with a double-blind, placebo-controlled, randomized drug withdrawal period of LUM001, an apical sodium-dependent bile acid transporter inhibitor (ASBTi), in patients with alagille syndrome: study LUM001-304; statistical analysis plan [unveröffentlicht]. 17.03.2020.
28. **Mohammad S, Kaur E, Aguirre VP, Varni JW, Alonso EM.** Health-related quality of life in infants with chronic liver disease. *J Pediatr Gastroenterol Nutr* 2016;62(5):751-756.
29. **Nixon A, Doll H, Kerr C, Burge R, Naegeli AN.** Interpreting change from patient reported outcome (PRO) endpoints: patient global ratings of concept versus patient global ratings of change, a case study among osteoporosis patients. *Health Qual Life Outcomes* 2016;14:25.
30. **Norman GR, Stratford P, Regehr G.** Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50(8):869-879.
31. **Shneider BL, Kamath BM, Magee JC, Goodrich NP, Loomes KM, Ye W, et al.** Use of funded multicenter prospective longitudinal databases to inform clinical trials in rare diseases: examination of cholestatic liver disease in Alagille syndrome. *Hepatol Commun* 2022;6(8):1910-1921.
32. **The Hospital for Sick Children (SickKids).** The GALA study [online]. Toronto (CAN): The Hospital for Sick Children (SickKids); 2022. [Zugriff: 07.03.2023]. URL: <https://www.galastudy.com/>.
33. **Vandriel SM, Li LT, She H, Wang JS, Gilbert MA, Jankowska I, et al.** Natural history of liver disease in a large international cohort of children with Alagille syndrome: Results from the GALA study. *Hepatology* 2023;77(2):512-529.
34. **Varni JW, Limbers CA, Neighbors K, Schulz K, Lieu JE, Heffer RW, et al.** The PedsQL Infant Scales: feasibility, internal consistency reliability, and validity in healthy and ill infants. *Qual Life Res* 2011;20(1):45-55.
35. **Whittington PF, Whittington GL.** Partial external diversion of bile for the treatment of intractable pruritus associated with intrahepatic cholestasis. *Gastroenterology* 1988;95(1):130-136.
36. **World Health Organization (WHO).** Child growth standards [online]. Genf (SUI): WHO. [Zugriff: 05.04.2023]. URL: <https://www.who.int/tools/child-growth-standards>.

## Anhang

### Diagnosekriterien des ALGS

Tabelle 37: Diagnostische Kriterien des ALGS

| ALGS-Familien-<br>vorgeschichte <sup>1)</sup> | Mangel an<br>Gallengängen         | JAGGED1- oder<br>NOTCH2-Mutation  | Für die Diagnose<br>benötigte klinische<br>Hauptkriterien <sup>2)</sup> |
|---|-----------------------------------|-----------------------------------|---|
| Vorhanden oder nicht<br>vorhanden             | Vorhanden                         | Identifiziert                     | Jegliches oder kein<br>Merkmal  |
| Keine (Proband)                               | Nicht vorhanden oder<br>unbekannt | Identifiziert                     | Ein oder mehrere<br>Merkmale  |
| Keine (Proband)                               | Vorhanden                         | Nicht identifiziert <sup>3)</sup> | Drei oder mehr<br>Merkmale  |
| Keine (Proband)                               | Nicht vorhanden oder<br>unbekannt | Nicht identifiziert <sup>3)</sup> | Vier oder mehr<br>Merkmale  |
| Vorhanden                                     | Nicht vorhanden oder<br>unbekannt | Identifiziert                     | Jegliches oder kein<br>Merkmal  |
| Vorhanden                                     | Vorhanden                         | Nicht identifiziert <sup>3)</sup> | Ein oder mehrere<br>Merkmale  |
| Vorhanden                                     | Nicht vorhanden oder<br>unbekannt | Nicht identifiziert <sup>3)</sup> | Zwei oder mehr<br>Merkmale  |

<sup>1)</sup> ALGS bei einer/einem Verwandten ersten Grades vorhanden.

<sup>2)</sup> Klinische Hauptkriterien/-merkmale für ALGS umfassen Cholestase, durchgängige Beteiligung des Herzens, der Nieren, der Gefäße, der Augen, des Skeletts oder charakteristische "Alagille"-Fazies.

<sup>3)</sup> Bei Screening nicht identifiziert oder nicht untersucht.

Abkürzungen: ALGS: Alagille-Syndrom.

### s-BA-Spiegel (Primärer Endpunkt der Studie ICONIC)

In Tabelle 38 sind die Ergebnisse des primären Endpunkts, Veränderung des Nüchtern-sBA-Spiegels zu Woche 22 gegenüber Woche 18 bei Personen, die zuvor auf die Maralixibat-Behandlung angesprochen haben, dargestellt. Da der sBA-Spiegel als Laborparameter als nicht patientenrelevant eingestuft wird, erfolgt die Darstellung lediglich ergänzend. Kritisch anzumerken ist, dass es sich bei der sogenannten mITT-Population um eine stark reduzierte Subpopulation aller Personen der Studie ICONIC handelt, die zu Woche 12 oder 18 eine Reduktion  $\geq 50\%$  des sBA-Spiegels im Vergleich zu Baseline erreichten (5 Personen (38 %) der Maralixibat-Gruppe, 10 Personen (63 %) der Placebo-Gruppe). Dieses Vorgehen führt auch zu einem Bruch der Randomisierung, sodass doppelt so viele Placebo-behandelte Personen in die Analyse eingehen als Personen, die weiterhin Maralixibat in der RW-Phase erhielten.

Unter Maralixibat zeigte sich zu Woche 22 eine weitere Reduktion des sBA-Spiegels im Vergleich zu Woche 18 mit einem LS MW von -21,7. Dagegen stieg der Wert in der Placebo-Gruppe im LS MW deutlich um 95,6 an. Der Effekt ist statistisch signifikant.

Tabelle 38: Veränderung des sBA-Werts in der RW-Phase; Studie ICONIC, mITT-Population

| Studie ICONIC<br>sBA-Wert                               | Maralixibat<br>N = 5 <sup>1)</sup> | Placebo<br>N = 10 <sup>1)</sup> |
|---|------------------------------------|---------------------------------|
| Woche 18<br>LS MW (SE)                                  | 100,22 (24,71)                     | 132,13 (17,40)                  |
| Woche 22<br>LS MW (SE)                                  | 68,83 (49,59)                      | 232,50 (34,91)                  |
| <b>Veränderung zu Woche 22 gegenüber Woche 18</b>       |                                    |                                 |
| LS MW (SE) <sup>2)</sup>                                | -21,73 (43,13)                     | 95,55 (30,49)                   |
| LS-Mittelwertdifferenz [95%-KI]<br>p-Wert <sup>2)</sup> | -117,28 [-232,38; -2,18]<br>0,0464 |                                 |

<sup>1)</sup> mITT-Population: Alle eingeschlossenen Personen, die die Studienmedikation bis Woche 18 erhalten haben und eine Reduktion von  $\geq 50\%$  im sBA-Wert zu Woche 12 oder 18 erreichten.

<sup>2)</sup> ANCOVA-Modell mit Behandlungsgruppe als festen Effekt und Baseline-Wert als Kovariate.

Abkürzungen: ANCOVA: Analysis of Covariance; KI: Konfidenzintervall; LS: Least Squares; mITT: Modifizierte Intention-to-Treat; MW: Mittelwert; sBA: Serumgallensäure; SE: Standardfehler.

### Ergänzende Auswertungen zum ItchRO

Tabelle 39: Anteil der Tage bzw. Personen mit einem Wert  $\leq 1$  Punkt im ItchRO in der Erhebung am Morgen; Studie ICONIC

| Studie ICONIC<br>ItchRO, Erhebung am Morgen                        | Gesamt<br>ItchRO(Obs)<br>N = 31 | Gesamt<br>ItchRO(Pt)<br>N = 16 <sup>1)</sup> |
|--|---------------------------------|--|
| <b>Screening</b>   |                                 |  |
| Gesamtzahl der Tage (= im Beobachtungszeitraum)                    | k. A.                           | k. A.  |
| Gesamtzahl der Tage mit Erhebungen, N <sup>2)</sup>                | 786                             | 364  |
| Tage mit $\leq 1$ Punkt, n (%)                                     | 47 (6)                          | 21 (6)                                       |
| Anteil Personen mit wöchentlichem Durchschnittswert $\leq 1$ Punkt | k. A.                           | k. A.  |
| <b>Tag 1 bis Woche 6</b>   |                                 |  |
| Gesamtzahl der Tage (= im Beobachtungszeitraum)                    | k. A.                           | k. A.  |
| Gesamtzahl der Tage mit Erhebungen, N <sup>2)</sup>                | 1289                            | 584  |
| Tage mit $\leq 1$ Punkt, n (%)                                     | 296 (23)                        | 237 (41)                                     |
| Anteil Personen mit wöchentlichem Durchschnittswert $\leq 1$ Punkt | k. A.                           | k. A.  |
| <b>Woche &gt; 6 bis 12</b>   |                                 |  |
| Gesamtzahl der Tage (= im Beobachtungszeitraum)                    | k. A.                           | k. A.  |
| Gesamtzahl der Tage mit Erhebungen, N <sup>2)</sup>                | 1076                            | 518  |
| Tage mit $\leq 1$ Punkt, n (%)                                     | 529 (50)                        | 360 (70)                                     |
| Anteil Personen mit wöchentlichem Durchschnittswert $\leq 1$ Punkt | k. A.                           | k. A.  |
| <b>Woche &gt; 12 bis 18</b>  |                                 |  |
| Gesamtzahl der Tage (= im Beobachtungszeitraum)                    | k. A.                           | k. A.  |
| Gesamtzahl der Tage mit Erhebungen, N <sup>2)</sup>                | 1262                            | 595  |
| Tage mit $\leq 1$ Punkt, n (%)                                     | 752 (60)                        | 479 (81)                                     |
| Anteil Personen mit wöchentlichem Durchschnittswert $\leq 1$ Punkt | k. A.                           | k. A.  |

| <b>Studie ICONIC<br/>ItchRO, Erhebung am Morgen</b>           | <b>Gesamt<br/>ItchRO(Obs)<br/>N = 31</b> | <b>Gesamt<br/>ItchRO(Pt)<br/>N =16<sup>1)</sup></b> |
|---|--|---|
| <b>Woche &gt; 22 bis 48</b>                                   |  |   |
| Gesamtzahl der Tage (=im Beobachtungszeitraum)                | k. A.                                    | k. A.   |
| Gesamtzahl der Tage mit Erhebungen, N <sup>2)</sup>           | 4618                                     | 2376  |
| Tage mit ≤ 1 Punkt, n (%)                                     | 3043 (66)                                | 2017 (85)   |
| Anteil Personen mit wöchentlichem Durchschnittswert ≤ 1 Punkt | k. A.                                    | k. A.   |

<sup>1)</sup> Eine Beantwortung war ab einschließlich 5 Jahren möglich, wobei die Subpopulation der ≥ 5-Jährigen in der Studie ICONIC aus 16 Personen bestand. Die Anzahl der Personen, bei denen der ItchRO(Pt) tatsächlich eingesetzt wurde, ist nicht beschrieben.

<sup>2)</sup> Anzahl der Erhebungen in der Bezugsperiode. Die maximal mögliche Anzahl an Erhebungen entspricht der kumulierten Anzahl der Tage, in denen sich alle Patientinnen und Patienten im Bezugszeitraum befanden.

Abkürzungen: ItchRO: Itch Reported Outcome; ItchRO(Obs): Itch Reported Outcome (Observer); ItchRO(Pt): Itch Reported Outcome (Patient); k. A.: keine Angabe.