

# 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**

**Neues Anwendungsgebiet: Progressive familiäre  
intrahepatische Cholestase (PFIC) bei Personen ab dem Alter  
von 3 Monaten**

Datum der Veröffentlichung: 1. November 2024

Version 1.0

## Inhaltsverzeichnis

Tabellenverzeichnis .....	3
Abkürzungsverzeichnis .....	5
Hintergrund.....	7
1 Fragestellung.....	8
2 Auswahl und Methodik der relevanten Studien .....	9
2.1 Studienbasis für die Nutzenbewertung .....	9
2.2 Beschreibung der eingeschlossenen Studien .....	10
2.2.1 Studie MARCH-PFIC.....	10
2.2.2 Studie RISE .....	16
2.3 Endpunkte .....	19
2.3.1 Mortalität.....	19
2.3.2 Morbidität.....	20
2.3.3 Lebensqualität .....	31
2.3.4 Sicherheit .....	32
2.3.5 Erhebungszeitpunkte .....	35
2.4 Statistische Methoden.....	37
2.5 Verzerrungspotential auf Studien- und Endpunktebene .....	40
3 Ergebnisse der eingeschlossenen Studien.....	42
3.1 Studiencharakteristika und Studienmedikation .....	42
3.2 Mortalität .....	49
3.3 Morbidität .....	49
3.4 Lebensqualität.....	54
3.5 Sicherheit .....	54
4 Diskussion der Methodik und Ergebnisse.....	59
4.1 Zulassungsstatus und Zulassungspopulation von Maralixibat.....	59
4.2 Design und Methodik der Studie.....	60
4.3 Mortalität .....	63
4.4 Morbidität .....	63
4.5 Lebensqualität.....	65
4.6 Sicherheit .....	65
5 Anforderung an eine qualitätsgesicherte Anwendung .....	67
6 Zusammenfassung der Nutzenbewertung .....	68
Referenzen .....	72
Anhang .....	74

## Tabellenverzeichnis

Tabelle 1: Übersicht über die Studienbasis .....	9
Tabelle 2: Charakterisierung der Studie MARCH-PFIC.....	10
Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie MARCH-PFIC .....	14
Tabelle 4: Charakterisierung der Studie MARCH-PFIC.....	15
Tabelle 5: Charakterisierung der Studie RISE .....	16
Tabelle 6: Charakterisierung der Intervention in der Studie RISE.....	18
Tabelle 7: Zusammenfassung der Endpunktbewertung der Studien MARCH-PFIC und RISE ....	19
Tabelle 8: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie MARCH-PFIC .....	35
Tabelle 9: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie RISE.....	36
Tabelle 10: Verzerrungspotential der Studie MARCH-PFIC.....	40
Tabelle 11: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie MARCH-PFIC; PFIC-Population.....	41
Tabelle 12: Allgemeine Angaben; Studie MARCH-PFIC (finaler Datenschnitt).....	42
Tabelle 13: Allgemeine Angaben; Studie RISE (Interim-Datenschnitt: 23.06.2022) .....	43
Tabelle 14: Charakterisierung der Studienpopulationen der Studien MARCH-PFIC und RISE ....	45
Tabelle 15: Responderanalyse zur Verbesserung im ItchRO(Obs) in der Erhebung am Morgen; Studie MARCH-PFIC, ITT-Population.....	50
Tabelle 16: Mittlere Veränderung im Schweregrad des ItchRO(Obs) zu Woche 15–26 gegenüber Baseline; Studie MARCH-PFIC, ITT-Population (ergänzend dargestellt).....	50
Tabelle 17: Veränderung des Scores der „Klinischen Kratzskala“ im Vergleich zu Baseline; Studie RISE, Sicherheitspopulation .....	52
Tabelle 18: Körperliche Entwicklung anhand Körpergröße (z-Score) und Körpergewicht (z-Score) zu Woche 18–26 gegenüber Baseline; Studie MARCH-PFIC, ITT-Population.....	53
Tabelle 19: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase; Studie MARCH-PFIC, Sicherheitspopulation .....	55
Tabelle 20: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Behandlungsphase; Studie RISE, Sicherheitspopulation .....	55
Tabelle 21: UE mit Inzidenz $\geq 10\%$ ; Studie MARCH-PFIC, Sicherheitspopulation .....	56
Tabelle 22: UE mit Inzidenz $\geq 10\%$ und aufgetreten bei mind. 2 Personen; Studie RISE, Sicherheitspopulation .....	57
Tabelle 23: UE von besonderem Interesse; Studie MARCH-PFIC, Sicherheitspopulation.....	58
Tabelle 24: Zusammenfassende Darstellung der Ergebnisse der Studie MARCH-PFIC .....	69
Tabelle 25: Zusammenfassende Darstellung der Ergebnisse der Studie RISE.....	71
Tabelle 26: Veränderung der sBA zu Woche 18–26 gegenüber Baseline; Studie MARCH-PFIC, ITT-Population.....	74

Tabelle 27: Responderanalysen zur Verbesserung (Reduktion) der sBA; Studie MARCH-PFIC, ITT-Population.....	74
Tabelle 28: Veränderung der sBA zu Woche 13 gegenüber Baseline; Studie RISE, Sicherheitspopulation .....	75

## Abkürzungsverzeichnis

ALGS	Alagille-Syndrom
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
BID	zweimal täglich (bis in die)
BSEP	Gallensalz-Exportpumpe (Bile Salt Export Pump)
CIS	Caregiver Impression of Severity
CMH	Cochran-Mantel-Haenszel
CTCAE	Common Terminology Criteria for Adverse Events
EDQ	Exploratory Diary Questionnaire
EPAR	European Public Assessment Report
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
HCC	Hepatozelluläres Karzinom (Hepatocellular Carcinoma)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ItchRO	Itch Reported Outcome
ITT	Intention to Treat
KI	Konfidenzintervall
LLT	Lowest Level Term
LS	Kleinste Quadrate (Least Squares)
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measures
MWD	Mittelwertdifferenz
N	Anzahl
nt-	non-truncated (nicht-trunkiert)
(Obs)	Observer Version
PEBD	Partielle externe Gallengangsableitung (Partial External Biliary Diversion)
PedsQL	Pediatric Quality of Life Inventory
PFIC	Progressive familiäre intrahepatische Cholestase
PIS	Patient Impression of Severity
PT	Preferred Term
(Pt)	Patient Version
pU	pharmazeutischer Unternehmer
QD	einmal täglich (quaque die)
RR	Relatives Risiko
SAP	Statistischer Analyseplan
sBA	Serum-Gallensäure (Serum Bile Acid)
SD	Standardabweichung
SGB	Sozialgesetzbuch
SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis

t-	truncated (trunkiert)
UDCA	Ursodeoxycholsäure (Ursodeoxycholic Acid)
UE	Unerwünschtes Ereignis
VerfO	Verfahrensordnung des G-BA

## Hintergrund

Maralixibat 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 29. Oktober 2024 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 29. Juli 2024 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 1. November 2024 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. Die Bewertung bezieht sich auf folgende Indikation [6]:

Progressive familiäre intrahepatische Cholestase (PFIC) bei Patienten ab dem Alter von 3 Monaten.

Gemäß Fachinformation beträgt die empfohlene Zieldosis 570 µg/kg zweimal täglich (BID). Die Anfangsdosis beträgt 285 µg/kg einmal täglich (QD) und kann nach ein bis zwei Wochen auf 285 µg/kg BID erhöht werden. Eine weitere Auftitration auf 570 µg/kg BID kann nach weiteren ein bis zwei Wochen erfolgen, falls klinisch angezeigt und je nach Verträglichkeit. Die Dosis von Maralixibat von 570 µg/kg BID als freie Base entspricht der Dosis von Maralixibat-Chlorid von 600 µg/kg BID.

Weiterhin wirkt Maralixibat durch die Hemmung des ilealen Gallensäuretransporters (IBAT) und die Unterbrechung des enterohepatischen Kreislaufs der Gallensäuren. Daher können Erkrankungen, Arzneimittel oder chirurgische Eingriffe, die entweder die gastrointestinale Motilität oder die enterohepatische Zirkulation von Gallensäuren, einschließlich des Gallensalztransports in die kleinsten Gallengänge, beeinträchtigen, die Wirksamkeit von Maralixibat vermindern.

Aus diesem Grund ist bei Patientinnen und Patienten mit PFIC2, bei denen das Protein Gallensalz-Exportpumpe (BSEP) vollständig fehlt oder nicht funktioniert (d. h. bei Patientinnen und Patienten mit dem BSEP3-Subtyp von PFIC2), kein Ansprechen auf Maralixibat zu erwarten.

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studiennummer (Studienname)	Studie mit Dossier vom pU eingereicht	Studie vom pU als relevant für die Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzen- bewertung	Ausschlussgrund
<b>Studie zum Wirkstoff</b>				
MRX-502 <sup>1)</sup> (MARCH-PFIC)	Ja	Ja	Ja	-
<b>Studie zu externen Kontrollen</b>				
MRX-801 <sup>2)</sup> (RISE)	Ja	Ja	Ja	-

<sup>1)</sup> Zulassungsrelevante Studie gemäß EPAR.

<sup>2)</sup> Im Rahmen des Zulassungsverfahrens bei der EMA ergänzend zu den Zulassungsstudien berücksichtigt.

Abkürzungen: EMA: European Medicines Agency; EPAR: European Public Assessment Report; pU: pharmazeutischer Unternehmer.

Die für die Nutzenbewertung herangezogenen Studien entsprechen den Studien, auf denen die Einschätzungen des pU zum Zusatznutzen basieren.

#### Zur Nutzenbewertung für Maralixibat herangezogene Unterlagen

- Dossier des pU zu Maralixibat [17]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [5]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der pivotalen Studie MRX-502 (MARCH-PFIC) [14,15,16] und der Studie MRX-801 (RISE) [11,12,13]
- Produktinformation zu Maralixibat [6]
- Publikationen zur pivotalen Studie MARCH-PFIC [10]

## 2.2 Beschreibung der eingeschlossenen Studien

Die Nutzenbewertung von Maralixibat im vorliegenden Anwendungsgebiet basiert auf der Zulassungsstudie MRX-502 (MARCH-PFIC) und der Studie MRX-801 (RISE). Die Studien und die Intervention werden in den Tabellen 2 bis 6 charakterisiert.

### 2.2.1 Studie MARCH-PFIC

Tabelle 2: Charakterisierung der Studie MARCH-PFIC

Charakteristikum	Beschreibung
<b>Design und Studienablauf</b>	<p>MARCH-PFIC ist eine multinationale, randomisierte, doppelblinde, placebokontrollierte Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit von Maralixibat bei Patientinnen und Patienten mit PFIC ab einem Alter von 12 Monaten bis &lt; 18 Jahren. Die Randomisierung erfolgte im Verhältnis 1:1 stratifiziert nach definierten Kohorten (s. u.) auf die Behandlungsarme Maralixibat und Placebo.</p> <p>Die Studie ist in folgende Abschnitte unterteilt:</p> <ul style="list-style-type: none"> <li>• Screening (bis zu 6 Wochen)</li> <li>• Dosisescalation (4–6 Wochen)</li> <li>• Erhaltungsphase (20–22 Wochen)</li> <li>• Sicherheitsnachbeobachtung (7 Tage)<sup>1)</sup></li> </ul> <p>Am Ende der 26-wöchigen doppelblinden Behandlungsphase haben die Patientinnen und Patienten die Möglichkeit, an der Extensionsstudie MRX-503 teilzunehmen. Wird das Studienmedikament vorzeitig und dauerhaft abgesetzt, erfolgt ein vorzeitiger Studienabbruch durch die Testperson inklusive Abschlussvisite. Für Personen, die nicht an der Extensionsstudie teilnehmen, sind jährliche Follow-up-Visiten geplant.</p> <p><u>Anhand der Ein- und Ausschlusskriterien waren 2 separate Kohorten definiert:</u></p> <p>a) Primäre Kohorte (randomisiert im Verhältnis 1:1 auf Maralixibat oder Placebo):</p> <ul style="list-style-type: none"> <li>○ Personen mit einer biallelischen erkrankungsverursachenden Variante im PFIC2 (ABCB11-Gen).</li> <li>○ Ausgenommen Personen mit t-PFIC2<sup>2)</sup> (BSEP3 im ABCB11-Gen, kompletter Verlust der BSEP-Funktion), Heterozygotie, niedrigem oder fluktuierendem sBA oder früherer Operation zur Behandlung der PFIC.</li> </ul> <p>b) Ergänzende Kohorte (randomisiert im Verhältnis 1:1 auf Maralixibat oder Placebo, stratifiziert nach PFIC1, PFIC3 und allen weiteren Studienteilnehmenden der Kohorte):</p> <ul style="list-style-type: none"> <li>○ Personen mit t-PFIC2.</li> <li>○ Personen mit einer biallelischen erkrankungsverursachenden Variante im PFIC1 (ATP8B1-Gen), PFIC3 (ABCB4-Gen) oder PFIC4 (TJP2-Gen).</li> <li>○ Personen mit PFIC-Phänotyp ohne bekannte Mutation, Heterozygotie mit anderen bekannten Mutationen oder mit intermittierender Cholestase, manifestiert durch fluktuierende sBA-Level.</li> <li>○ Personen mit PFIC nach interner oder externer Gallengang-Operation oder bei denen die Operation rückgängig gemacht wurde.</li> </ul> <p><u>Aus den Kohorten wurden unterschiedliche Populationen für die Auswertung gebildet:</u></p> <ol style="list-style-type: none"> <li>1. Primäre-Population: Alle Personen der primären Kohorte (nt-PFIC2)<sup>3)</sup></li> <li>2. PFIC-Population: <ul style="list-style-type: none"> <li>○ Teilnehmende mit PFIC1, nt-PFIC2, PFIC3, PFIC4 (TJP2-Gen) und PFIC6 (MYO5B-Gen) (alle bekannten etablierten Genotypen außer Personen mit t-PFIC2).</li> <li>○ Ausgenommen Personen mit Heterozygotie, niedrigem oder fluktuierendem sBA oder früherer Operation zur Behandlung der PFIC.</li> </ul> </li> <li>3. Gesamtpopulation: Alle Studienteilnehmenden; inklusive Personen, die nicht die Kriterien für die Primäre- oder die PFIC-Population erfüllten.</li> </ol>

Charakteristikum	Beschreibung
	<p>Die Studie ist abgeschlossen (Studienbericht: 18.05.2023).</p> <p>Die auf nt-PFIC2 beschränkte Primäre-Population stellt nur einen Teil des zugrundeliegenden Anwendungsgebiets dar und wird im Folgenden nicht weiter betrachtet.</p>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Körpergewicht <math>\geq 5,0</math> kg, im Alter <math>\geq 12</math> Monate und <math>&lt; 18</math> Jahre zu Baseline.</li> <li>• Cholestase, manifestiert durch Gesamt-sBA <math>\geq 3 \times</math> ULN (primäre Kohorte).</li> <li>• Mittlerer morgendlicher ItchRO(Obs)-Score <math>\geq 1,5</math> für 4 aufeinanderfolgende Wochen in der Screeningphase bis zur Baseline-Visite (Visite 1).</li> <li>• Verfügbare Daten von <math>\geq 21</math> validen morgendlichen ItchRO(Obs)-Einträgen während 4 konsekutiven Wochen in der Screeningphase bis zur Baseline-Visite.<sup>4)</sup></li> <li>• Diagnose der PFIC basierend auf: <ul style="list-style-type: none"> <li>○ Chronische Cholestase, manifestiert als persistierender (<math>&gt; 6</math> Monate) Pruritus zusammen mit biochemischen Abnormalitäten und/oder pathologischer Evidenz für eine progressive Lebererkrankung und</li> <li>○ Primäre Kohorte: einem genetischen Testresultat PFIC2, basierend auf Standard-Genotypisierung.</li> <li>○ Ergänzende Kohorte: <ul style="list-style-type: none"> <li>i. einem genetischen Testresultat PFIC1, PFIC3 oder PFIC4, basierend auf Standard-Genotypisierung;</li> <li>ii. PFIC-Phänotyp ohne bekannte Mutation oder mit anderen bekannten Mutationen oder mit intermittierender Cholestase, manifestiert durch fluktuierende sBA-Level;</li> <li>iii. PFIC nach interner oder externer Gallengang-Operation oder, wenn die Operation rückgängig gemacht wurde.</li> </ul> </li> </ul> </li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Prognostizierte komplette Abwesenheit der BSEP-Funktion basierend auf dem Typ der Mutation des ABCB11-Gens (t-PFIC2; nur primäre Kohorte).<sup>5)</sup></li> <li>• Rekurrende intrahepatische Cholestase, indiziert durch frühere sBA-Level <math>&lt; 3 \times</math> ULN oder intermittierenden Pruritus (nur primäre Kohorte).<sup>5)</sup></li> <li>• Aktuelle oder frühere (<math>&lt; 1</math> Jahr) atopische Dermatitis oder andere nicht-cholestatische Erkrankungen mit assoziiertem Pruritus.</li> <li>• Frühere operative Unterbrechung der enterohepatischen Zirkulation (nur primäre Kohorte).<sup>5)</sup></li> <li>• Chronische Diarrhö, die eine ständige intravenöse Flüssigkeits- oder Nahrungszufuhr erfordert.</li> <li>• Vorherige oder Bedarf einer imminenten Lebertransplantation.</li> <li>• Dekompensierte Zirrhose, Vorgeschichte oder Vorhandensein von klinisch signifikantem Aszites, Varizenblutung und/oder Enzephalopathie.</li> <li>• Vorgeschichte oder Vorhandensein einer anderen begleitenden Lebererkrankung.</li> <li>• Krebserkrankungen (außer In-situ-Karzinome) innerhalb von 5 Jahren vor dem Screening.</li> <li>• Verabreichung von Gallensäure- oder Lipid-bindenden Granulaten, Phenylbutyrat oder jedweder anderer Prüfpräparate (Arzneimittel, Biologika, Medizinprodukte) während der Screeningphase.</li> </ul>

Charakteristikum	Beschreibung
<b>Intervention und Zahl der Patientinnen und Patienten</b>	<p><b>Gesamtpopulation</b> Gescreent gesamt: N = 125 Randomisiert: N = 93 Intervention: N = 47 Kontrolle: N = 46</p> <p><b>Primäre-Population<sup>6)</sup></b> Intervention: N = 14 Kontrolle: N = 17</p> <p><b>PFIC-Population<sup>6)</sup></b> Intervention: N = 33 Kontrolle: N = 31</p>
<b>Ort und Zeitraum der Durchführung; Datenschnitte</b>	<p><b>Studienzentren</b> 29 Zentren in 16 Ländern (Argentinien, Belgien, Brasilien, Deutschland, Frankreich, Italien, Kanada, Kolumbien, Libanon, Mexiko, Österreich, Polen, Singapur, Türkei, Vereinigtes Königreich, USA).</p> <p><b>Studienzeitraum</b> Erste Person, erste Visite: 09.07.2019 Letzte Person, letzte Visite: 01.09.2022</p> <p><b>Datenschnitt: 14.10.2022</b></p>
<b>Endpunkte gemäß Studienprotokoll</b>	<p><b>Primärer Endpunkt</b> Veränderung der Pruritus-Schwere gemessen mittels ItchRO(Obs) (Primäre-Population)<sup>7)</sup></p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Gesamt-sBA</li> <li>• Veränderung der Pruritus-Schwere gemessen mittels ItchRO(Obs) (PFIC-Population)</li> <li>• Pruritus-Responder gemessen mittels ItchRO(Obs)</li> <li>• sBA-Responder</li> <li>• Sicherheit und Verträglichkeit (UE), Pharmakokinetik</li> </ul> <p><b>Explorative Endpunkte u. a.</b></p> <ul style="list-style-type: none"> <li>• Leberassoziierte Laborwerte</li> <li>• Leber-Biochemie</li> <li>• Nichtinvasive Fibrosemarker: FIB-4, APRI, PELD-Score, MELD-Score</li> <li>• Marker für Gallensäuremetabolismus</li> <li>• Pruritus gemessen mittels ItchRO(Obs) und ItchRO(Pt)</li> <li>• Klinische Kratzskala (CSS)</li> <li>• Ansprechraten für den Schweregrad von Pruritus und Gesamt-sBA</li> <li>• PedsQL inklusive „Family Impact Module“ und „Multidimensional Fatigue Scale“</li> <li>• Schlafstörungen mittels EDQ</li> <li>• Wachstum anhand Körpergröße und Körpergewicht</li> <li>• Zeit bis zu leberassoziierten Ereignissen</li> </ul>

<sup>1)</sup> Gilt für Personen, die die Studie vorzeitig abbrechen sowie für jene, die nicht in die Extensionsstudie MRX-503 wechseln.

<sup>2)</sup> t = truncated; das BSEP-Protein ist verkürzt, was seine Funktion vollständig verhindert. In dieser Population ist gemäß Fachinformation kein Ansprechen auf Maralixibat zu erwarten.

<sup>3)</sup> nt = non-truncated; das BSEP-Protein ist nicht verkürzt.

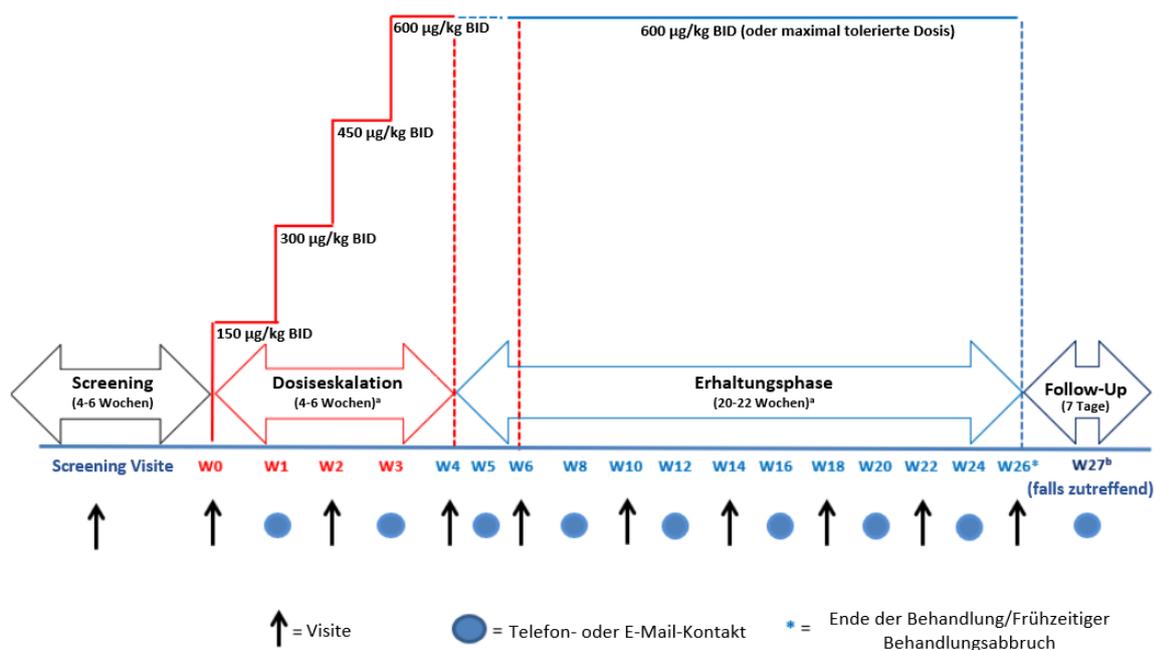
<sup>4)</sup> Valide = beantwortet, jedoch nicht mit „Ich weiß nicht“; ≤ 2 nicht valide Berichte während der 7 Tage vor Randomisierung.

<sup>5)</sup> Personen konnten an der Studie als Teil der ergänzenden Kohorte teilnehmen.

<sup>6)</sup> Es besteht eine Überschneidung zwischen der Primäre- und der PFIC-Population. Alle Teilnehmenden der Primäre-Population sind auch in der PFIC-Population enthalten (siehe oben).

<sup>7)</sup> In der Studie MARCH-PFIC wurden zudem der PIC und CIC erhoben. Diese waren jedoch nicht als Endpunkte definiert und es liegen keine Ergebnisse vor. Nach Angaben des pU sollten sie zur Validierung des ItchRO beitragen.

Abkürzungen: APRI: AST-to-Platelet Ratio Index; AST: Aspartat-Aminotransferase; BSEP: Gallensalz-Exportpumpe; CIC: Caregiver Impression of Change; CSS: Clinician Scratch Scale; EDQ: Exploratory Diary Questionnaire; FIB-4: Fibrosis-4-Index; ItchRO(Obs): Itch Reported Outcome (Observer Version); ItchRO(Pt): Itch Reported Outcome (Patient Version); MELD: Model for End-Stage Liver Disease; nt-PFIC: non-truncated PFIC; PedsQL: Pediatric Quality of Life Inventory; PELD: Pediatric End-Stage Liver Disease; PFIC: Progressive familiäre intrahepatische Cholestase; PIC: Patient Impression of Change; pU: pharmazeutischer Unternehmer; sBA: Serum-Gallensäure; t-PFIC: truncated PFIC; UE: Unerwünschtes Ereignis; ULN: Upper Limit of Normal.



<sup>a</sup> Dosisescalation je nach Verträglichkeit über einen Zeitraum von 4-6 Wochen; Stabile Dosis über 20-22 Wochen, in Abhängigkeit von der Dosisescalationsdauer.  
<sup>b</sup> Sicherheits-Follow-Up für Teilnehmende, die nicht in die Extensionsstudie übergehen (MRX-503).

Abbildung 1: Schematische Darstellung des Studienablaufs der Studie MARCH-PFIC

## Protokolländerungen

Es wurden 22 Änderungen des Originalprotokolls vom 14.03.2019 vorgenommen, davon 18 landes-spezifische Amendments. Da vor allem die landesübergreifenden Amendments relevante Änderungen umfassen, werden diese in Tabelle 3 dargestellt; davon wurden 3 nach Einschluss der ersten Person (09.07.2019) implementiert.

Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie MARCH-PFIC

Amendment	Wesentliche Änderungen
Amendment 2 22.11.2019 (Anzahl bis dahin eingeschlossene Personen: k. A.)	<ul style="list-style-type: none"> <li>• PIC und CIC ergänzt zur letzten Visite; „Klinische Kratzskala“ zum Screening ergänzt.</li> <li>• Ergänztes Ausschlusskriterium zu rekurrender intrahepatischer Cholestase zur Vermeidung des Einschlusses von Personen mit intermittierend spontaner Normalisierung des Pruritus und sBA oder Remission des Pruritus.</li> <li>• Ergänztes Ausschlusskriterium zum Ausschluss von Personen mit nicht-Cholestase bedingtem Pruritus.</li> <li>• Anpassung des Ausschlusskriteriums und Kriteriums zur vorzeitigen Beendigung der Studie von geplanter zu imminenter Lebertransplantation.</li> </ul>
Amendment 3 16.06.2020 (Anzahl bis dahin eingeschlossene Personen: k. A.)	<ul style="list-style-type: none"> <li>• Aktualisiertes Einschlusskriterium der Gesamt-sBA <math>&gt; 3 \times \text{ULN}</math> für die primäre Kohorte.</li> <li>• Klarstellung, dass zur Diagnose einer PFIC ein persistierender Pruritus <i>zusätzlich</i> zu biochemischen Aberrationen und/oder pathologischen Nachweisen einer progressiven Lebererkrankung vorliegen müssen.</li> </ul>
Amendment 4 10.05.2022 (Anzahl bis dahin eingeschlossene Personen: k. A.)	<ul style="list-style-type: none"> <li>• Aktualisierte sekundäre und explorative Studienziele und Endpunkte zur besseren Erhebung klinisch relevanter Messungen in der Primäre- und PFIC-Population.</li> <li>• Definition der Analysepopulationen mit nt-PFIC2-Personen als primäre Analysepopulation. Die PFIC-Population wird zudem für die sekundären und explorativen Endpunkte definiert. Zusätzlich wird für die Auswertung von explorativen Endpunkten auch die Gesamtpopulation definiert, welche alle Studienteilnehmenden enthält.</li> <li>• Überarbeitung der Auswertung zur Pruritus-Schwere von ItchRO(Obs) und ItchRO(Pt), zur zeitlichen Orientierung der Erhebung der Mittelwerte an der ersten Dosis der Studienmedikation sowie zur Bildung der Monatswerte.</li> </ul>

Abkürzungen: CIC: Caregiver Impression of Change, ItchRO(Obs): Itch Reported Outcome (Observer Version); ItchRO(Pt): Itch Reported Outcome (Patient Version); k. A.: keine Angabe; nt-PFIC: non-truncated PFIC; PFIC: Progressive familiäre intrahepatische Cholestase; PIC: Patient Impression of Change; sBA: Serum-Gallensäure; ULN: Upper Limit of Normal.

## Charakterisierung der Intervention

Tabelle 4: Charakterisierung der Studie MARCH-PFIC

Intervention	Kontrolle
<p><b>Maralixibat in einer oralen Lösung</b> Zweimal täglich; die erste Dosis sollte ca. 30 Minuten vor der ersten Mahlzeit des Tages, die zweite Dosis ca. 30 Minuten vor der Abendmahlzeit verabreicht werden.</p> <p><u>Dosiseskalationsperiode (Abschluss zwischen Woche 4 u. 6):</u></p> <ul style="list-style-type: none"> <li>• Dosis-Level 1: 150 µg/kg BID für mind. 1 Woche</li> <li>• Dosis-Level 2: 300 µg/kg BID für mind. 1 Woche</li> <li>• Dosis-Level 3: 450 µg/kg BID für mind. 1 Woche</li> </ul> <p><u>Erhaltungsphase (Beginn zwischen Woche 4 u. 6; Abschluss in Woche 26):</u></p> <ul style="list-style-type: none"> <li>• Stabile Dosis: 600 µg/kg BID bzw. maximal tolerierte Dosis</li> </ul>	<p><b>Placebo in einer oralen Lösung</b> Enthält alle Bestandteile des Prüfpräparats, mit Ausnahme des aktiven Wirkstoffs. Zweimal täglich; die erste Dosis sollte ca. 30 Minuten vor der ersten Mahlzeit des Tages, die zweite Dosis ca. 30 Minuten vor der Abendmahlzeit verabreicht werden.</p> <p>Verabreichung in Dosiseskalationsperiode und Erhaltungsphase wie Intervention.</p>
<p>Im Fall von Unverträglichkeiten (z. B. gastrointestinale Symptome) kann eine Dosisreduktion erfolgen; mind. 150 µg/kg BID sind jedoch erforderlich, um die Studie fortzusetzen.</p>	
<p><b>Nicht erlaubte Begleitmedikation<sup>1)</sup></b> Zu jeglichem Zeitpunkt während der Studie:</p> <ul style="list-style-type: none"> <li>• Gallensäure- oder Lipid-bindende Harze</li> <li>• Phenylbutyrat</li> </ul> <p><b>Erlaubte Begleitmedikation</b> Begleitmedikationen in einer stabil gehaltenen Dosis ab 30 Tage vor der Screening-Visite; ausgenommen Dosisadaptionen aufgrund von Gewichtsänderungen während der Studie:</p> <ul style="list-style-type: none"> <li>• Ursodeoxycholsäure oder Cholsäure</li> <li>• Rifampicin</li> <li>• Diphenhydramin</li> <li>• Hydroxyzin</li> <li>• Naltrexon</li> <li>• Naloxon</li> <li>• Phenobarbital</li> <li>• Antihistaminika</li> </ul>	

<sup>1)</sup> Chirurgische Interventionen sind bei den Begleitmedikationen nicht aufgeführt, allerdings ist der unmittelbare Bedarf einer Lebertransplantation, sowie eine Krankheitsprogression, die eine PEBD oder Lebertransplantation erforderlich macht, als Gründe für einen Studienabbruch gelistet.

Abkürzungen: BID: zweimal täglich; PEBD: Partielle externe Gallengangsableitung.

## 2.2.2 Studie RISE

Tabelle 5: Charakterisierung der Studie RISE

Charakteristikum	Beschreibung
<b>Design und Studienablauf</b>	<p>RISE ist eine offene, multizentrische Phase-II-Studie zur Untersuchung der Sicherheit und Verträglichkeit von Maralixibat 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>• Kernstudie: <ul style="list-style-type: none"> <li>○ Dosisescalation (Woche 1–6)</li> <li>○ Erhaltungsphase (Woche 7–13)</li> </ul> </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>Die Studie ist laufend (Interim-Studienbericht: 22.12.2022).</p> <p>Da für die vorliegende Nutzenbewertung ausschließlich Personen mit PFIC relevant sind, wird im Folgenden nur die PFIC-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</math> kg.</li> <li>• PFIC-Diagnose basierend auf genetischen Testergebnissen, die mit mind. einer krankheitsverursachenden, PFIC-assoziierten Mutation übereinstimmen, sowie mit klinischen oder biochemischen Hinweisen auf Cholestase.</li> <li>• Nachweis einer Cholestase (durch eines oder mehrere der folgenden Kriterien): <ul style="list-style-type: none"> <li>○ Gesamt-sBA <math>&gt; 2 \times</math> ULN für das betreffende Alter.</li> <li>○ Konjugiertes Bilirubin <math>&gt; 1</math> mg/dl.</li> <li>○ Mangel an fettlöslichen Vitaminen, der nicht durch andere Ursachen erklärbar ist.</li> <li>○ GGT <math>&gt; 3 \times</math> ULN für das betreffende Alter.</li> <li>○ Hartnäckiger Juckreiz (Pruritus), der nur durch eine Lebererkrankung erklärbar ist.</li> </ul> </li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Prognostizierte komplette Abwesenheit der BSEP-Funktion basierend auf dem Typ der Mutation des ABCB11-Gens (t-PFIC2).</li> <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.</li> <li>• ALT oder Gesamtbilirubin im Serum <math>&gt; 15 \times</math> 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>

Charakteristikum	Beschreibung
<b>Intervention und Zahl der Patientinnen und Patienten</b>	Gescreent gesamt: N = 4 Randomisiert/Eingeschlossen: N = 4 Intervention: N = 4
<b>Ort und Zeitraum der Durchführung; Datenschnitte</b>	<b>Studienzentren</b> 15 Studienzentren in 5 Ländern (USA, Großbritannien, Frankreich, Belgien, Polen). <b>Studienzeitraum</b> Studienstart: 09.09.2021 <sup>1)</sup> Laufende Studie; Studienabschluss für Dezember 2024 geplant <sup>1)</sup> . <b>Datenschnitt</b> Interim-Datenschnitt: 23.06.2022
<b>Endpunkte gemäß Studienprotokoll</b>	<b>Primärer Endpunkt</b> Sicherheit (UE) und Verträglichkeit <b>Sekundäre Endpunkte</b> <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> <b>Explorative Endpunkte</b> <ul style="list-style-type: none"> <li>• Pruritus mittels ItchRO</li> <li>• Pruritus mittels „Klinischer Kratzskala“</li> <li>• Wachstum (Größe, Gewicht, Körpermaße)</li> <li>• Nutzung des Gesundheitssystems (z. B. Hospitalisierungen)</li> </ul>

<sup>1)</sup> Gemäß Clinicaltrials.gov.

Abkürzungen: ALGS: Alagille-Syndrom; ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; BSEP: Gallensalz-Exportpumpe; GGT: Gamma-Glutamyltransferase; ItchRO: Itch Reported Outcome; PFIC: Progressive familiäre intrahepatische Cholestase; sBA: Serum-Gallensäure; t-PFIC: truncated PFIC; UE: Unerwünschtes Ereignis; ULN: Upper Limit of Normal.

### Protokolländerungen

Insgesamt wurden 11, teils landesspezifische, Änderungen am Protokoll vorgenommen. Davon wurden 6 Protokolländerungen nach dem Einschluss der ersten Person vorgenommen (09.09.2021), dabei 5 landesspezifische. Eine Auflistung der implementierten Anpassungen und Änderungen dieser Protokolländerungen konnte den Studienunterlagen nicht entnommen werden. Erst mit dem USA-spezifischen Protokollamendment 4 (14.01.2022) wurde die Erhebung des ItchRO(Obs) aufgenommen und darauffolgend auch im länderübergreifenden Amendment 5 (14.01.2022) festgehalten. Es ist daher unklar, ob für alle Studienteilnehmenden überhaupt Baseline-Daten zum ItchRO vorliegen.

## Charakterisierung der Intervention

Tabelle 6: Charakterisierung der Intervention in der Studie RISE

Intervention
<p><b>Maralixibat in einer oralen Lösung</b> Zweimal täglich; die erste Dosis sollte ca. 30 Minuten vor der ersten Mahlzeit des Tages, die zweite Dosis ca. 30 Minuten vor der Abendmahlzeit verabreicht werden.</p> <p><u>Dosiseskalationsperiode (Abschluss zwischen Woche 4 u. 6):</u></p> <ul style="list-style-type: none"> <li>• Dosis-Level 1: Maralixibat-Chlorid, 300 µg/kg QD für mind. 1 Woche</li> <li>• Dosis-Level 2: Maralixibat-Chlorid, 300 µg/kg BID für mind. 1 Woche</li> <li>• Dosis-Level 3: Maralixibat-Chlorid, 600 µg/kg BID für mind. 1 Woche</li> </ul> <p><u>Erhaltungsphase (Beginnt zwischen Woche 4 u. 6; Abschluss in Woche 13):</u></p> <ul style="list-style-type: none"> <li>• Stabile Dosis: MRX-Chlorid, 600 µg/kg BID</li> </ul> <p>Im Alter von &lt; 1 Monat beträgt die Dosis 75 µg/kg QD für die Dauer der Behandlung unterhalb dieser Altersgrenze.</p> <p>Im Fall von Unverträglichkeiten (z. B. gastrointestinale Symptome) kann eine Dosisreduktion erfolgen; mind. 150 µg/kg BID 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>• Präparate, die durch Alkoholdehydrogenasen metabolisiert werden</li> </ul>
<p><b>Erlaubte Begleitmedikation</b> Alle Behandlungen, die zum Behandlungsstandard zählen (z. B. Supplemente fettlöslicher Vitamine).</p>

Abkürzungen: BID: zweimal täglich; CTCAE: Common Terminology Criteria for Adverse Events; QD: einmal täglich; 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 MARCH-PFIC und RISE

Zusammenfassung der Endpunktbewertung	Kategorie	Studie		Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung	
		MARCH-PFIC	RISE			
Todesfälle	Mortalität	Ja	Ja	Ja	Ja	
Pruritus mittels <ul style="list-style-type: none"> <li>• ItchRO(Obs)<sup>1)</sup></li> <li>• ItchRO(Pt)</li> <li>• Eindruck d. Schwere (CIS/PIS)</li> <li>• EDQ(Obs)</li> <li>• Klinische Kratzskala</li> </ul>	Morbidity	Ja	Ja	Ja	Ja	
		Ja	Nein	Nein	Ja	Ja
		Ja	Nein	Ja	Nein	Nein
		Ja	Nein	Ja	Nein	Nein
		Ja	Ja	Ja	Ja	Ergänzend
Schlafstörungen mittels EDQ(Obs)		Ja	Nein	Ja	Ja	Ja
Körperliche Entwicklung		Ja	Ja	Ja	Ja	Ja
PedsQL-Fatigue	Ja	Nein	Ja	Ja	Ja	
sBA	Ja	Ja	Ja	Ja	Nein <sup>2)</sup>	
Leberassoziierte Ereignisse	Ja	Nein	Nein	Nein	Ergänzend	
PedsQL	Lebensqualität	Ja	Nein	Ja	Ja	
Unerwünschte Ereignisse <sup>3)</sup>	Sicherheit	Ja	Ja	Ja	Ja	

<sup>1)</sup> Primärer Endpunkt der Studie MARCH-PFIC.

<sup>2)</sup> Ergebnisse aus Transparenzgründen im Anhang dargestellt.

<sup>3)</sup> Primärer Endpunkt der Studie RISE.

Abkürzungen: CIS: Caregiver Impression of Severity; EDQ: Exploratory Diary Questionnaire; ItchRO(Obs): Itch Reported Outcome (Observer Version); ItchRO(Pt): Itch Reported Outcome (Patient Version); PedsQL: Pediatric Quality of Life Inventory; PIS: Patient Impression of Severity; pU: pharmazeutischer Unternehmer; sBA: Serum-Gallensäure.

### 2.3.1 Mortalität

#### Todesfälle

Der Endpunkt „Todesfälle“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Todesfälle wurden sowohl in der Studie MARCH-PFIC als auch in der Studie RISE 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 Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen.

### Validität

Gemäß Dossier wurden sowohl für die Studie MARCH-PFIC als auch für die Studie RISE nur Todesfälle aufgrund von UE erfasst. Laut Studienprotokoll und SAP sollten für MARCH-PFIC und RISE jedoch alle während der Studie aufgetretenen Todesfälle dokumentiert werden.

## **2.3.2 Morbidität**

### **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 MARCH-PFIC sowohl in einer betreuerberichteten (ItchRO(Obs)) als auch einer patientenberichteten (ItchRO(Pt)) Version zum Einsatz. Der ItchRO(Pt) wurde von allen Patientinnen und Patienten ab 9 Jahren (Alter zur Screening-Visite) beantwortet; Kinder unter 9 Jahren füllten den ItchRO(Pt) nicht aus. Im Folgenden sind mit der Bezeichnung „ItchRO“ immer beide Versionen des Instruments gemeint. Beide Versionen wurden zweimal täglich erhoben, morgens und abends. Auswertungen erfolgten auf Ebene der Einzelitems.

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 (Item 1), mit folgenden Antwortoptionen auf einer 5-Punkte-Likert-Skala:

- 0 („Es juckte mich nicht“)
- 1 („Es juckte mich ein bisschen“)
- 2 („Es juckte mich ziemlich“)
- 3 („Es juckte mich sehr stark“)
- 4 („Es juckte mich sehr, sehr stark“)

Zusätzlich wurde gefragt, ob der Juckreiz das Einschlafen erschwerte und ob der Juckreiz zum nächtlichen Aufwachen führte (jeweils mit „Ja“ oder „Nein“ zu beantworten; Items 2 und 3).

Schließlich wurde noch abgefragt, zu welchem Anteil der Nacht aufgrund des Juckens gerubbelt oder gekratzt wurde (Item 4), mit folgenden 5 Antwortoptionen:

- 0 („Zu keinem Zeitpunkt der Nacht“)
- 1 („Nur einen kleinen Teil der Nacht“)
- 2 („Die Hälfte der Nacht“)
- 3 („Die meiste Zeit der Nacht“)
- 4 („Die ganze Nacht“)

Am Abend erfolgte eine weitere Abfrage zum Juckreiz am Tag, mit denselben Items und Antwortoptionen der morgendlichen Erhebung. Falls die Betroffenen einen 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 2). Weiterhin sollte analog zur morgendlichen Abfrage angegeben werden, zu welchem Anteil des Tages aufgrund des Juckreizes gerubbelt oder gekratzt wurde (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 zusammenfassende Einstufung des Pruritus vornehmen (Item 1):

- 0 („Keine Symptome“)
- 1 („Mild“)
- 2 („Moderat“)
- 3 („Schwerwiegend“)
- 4 („Sehr schwerwiegend“)
- „Ich weiß es nicht“

Zudem 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 abgeben, wie viel Zeit das Kind mit Kratzen oder Rubbeln verbracht hat (Item 3):

- 0 („Gar nicht“)
- 1 („Nur einen kleinen Teil der Zeit“)
- 2 („Einige Zeit“)
- 3 („Die meiste Zeit“)
- 4 („Fast die gesamte Zeit / Konstant“)
- „Ich weiß nicht“

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. Für die Items 1 und 3 existieren die Antwortoptionen „Ich weiß nicht“ zum Opt-out, die als fehlende Werte gewertet werden. Falls möglich sollte der ItchRO(Obs) immer durch die gleiche betreuende Person ausgefüllt werden.

Die beiden Versionen des Instruments werden separat ausgewertet. Die primäre Auswertung erfolgt auf Basis der Veränderung des wöchentlichen, morgendlichen Durchschnittswerts des ItchRO(Obs) zwischen Baseline und Woche 15–26 in der Primäre-Population, basierend auf Item 1 (Pruritus-Schweregrad). Die Auswertung der Häufigkeit des Pruritus erfolgt anhand von Item 3 des ItchRO(Obs). Für den ItchRO(Pt) erfolgte die Auswertung des Schweregrads ebenfalls anhand Item 1 des Tagebuchs am Morgen und die Auswertung der Häufigkeit anhand Item 4.

Gemäß SAP sind folgende weitere Auswertungen der Pruritus-Werte prädefiniert:

- Veränderung des wöchentlichen, morgendlichen Durchschnittswerts des ItchRO(Obs) zwischen Baseline und Woche 15–26 in der PFIC-Population.
- Anteil der ItchRO(Obs) Responder zwischen Woche 15 und Woche 26.
- Anzahl der Tage mit einem morgendlichen ItchRO(Obs)-Schweregrad  $\leq 1$  auf Patientenebene für die letzten 84 Tage der Studie (Tag 99 bis einschließlich Tag 182).
- Anzahl und Anteil an Teilnehmenden mit morgendlichem ItchRO(Obs)-Schweregrad  $\leq 1$  für mehr als 50 % der Zeit zwischen Baseline und Woche 6, Woche 7–10, Woche 11–14, Woche 15–18, Woche 19–22, Woche 23–26 und Woche 15–26.
- Mittlere Veränderung zwischen Baseline und Woche 6, 10, 14, 18, 22 und 26, unter Nutzung des 6-wöchigen Durchschnitts in Woche 6 und nachfolgend des 4-wöchigen Durchschnitts, sowie des Durchschnitts von Woche 15 bis 26 kombiniert für:
  - den durchschnittlichen, morgendlichen ItchRO-Schweregrad
  - den durchschnittlichen, abendlichen ItchRO-Schweregrad
  - den durchschnittlichen höchsten täglichen ItchRO-Schweregrad
  - den durchschnittlichen, morgendlichen ItchRO-Häufigkeitswert
  - den durchschnittlichen, abendlichen ItchRO-Häufigkeitswert
  - den durchschnittlichen höchsten täglichen ItchRO-Häufigkeitswert

Eine Auswertung der dichotomen Items 2 und 3 des ItchRO(Pt) bezüglich der Auswirkungen des Pruritus auf das Ein- und das Durchschlafen erfolgte nicht.

Berechnung der Durchschnittswerte: Für jede Woche nach Baseline wird ein wöchentlicher Durchschnitt berechnet, wenn Daten für mindestens 4 Tage vorliegen. Die Auswertung erfolgt anhand von 4-wöchentlichen Durchschnittswerten, wobei zur Berechnung des Werts von Woche 15 bis 26 der Durchschnitt aus den 4-Wochen-Durchschnitten 15–18, 19–22, und 23–26 gebildet wird.

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

Als Responder werden Personen mit einer durchschnittlichen Reduktion im morgendlichen ItchRO um mindestens 1 Punkt im Vergleich zu Baseline ODER jene mit einem durchschnittlichen Schweregrad von  $\leq 1$  im jeweiligen Betrachtungszeitraum gewertet.

Im Fall der Studie RISE wurde der ItchRO(Obs) eingesetzt, jedoch ohne die Abfrage der Häufigkeit des Juckreizes. Primär sollte der wöchentliche durchschnittliche Schweregrad am Morgen berichtet und ausgewertet werden.

### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Es liegen ausschließlich Ergebnisse für die Items zu Pruritus-Schwere (Item 1) und -Häufigkeit (Item 3 bzw. Item 4) vor, eine Auswertung der weiteren Items war nicht geplant. Auf die durchgeführten Auswertungen wird detailliert in Kapitel 2.4 eingegangen.

### Patientenrelevanz

Pruritus zählt zu den häufigsten und belastenden Symptomen der PFIC [1,2,7]. Er kann zu Schlafstörungen, Reizbarkeit, offenen Wunden sowie Hautschäden führen und wird in der Literatur zudem als Grund für chirurgische Eingriffe oder die Notwendigkeit von Lebertransplantationen angegeben [18].

In der Studie MARCH-PFIC 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 klinischer Expertinnen und Experten und der U.S. Food and Drug Administration (FDA) [7]. Kern des Entwicklungsprozesses waren Interviews mit Kindern und Erwachsenen mit Alagille-Syndrom (ALGS) oder PFIC im Alter von 4 bis 30 Jahren sowie mit Betreuungspersonen von Patientinnen und Patienten im Alter von 5 Monaten bis 14 Jahren. Allerdings war in der Entwicklungsstudie lediglich 1 Person mit PFIC1 enthalten.

Die Validität des ItchRO wurde in einem vorherigen Nutzenbewertungsverfahren zu Maralixibat im Anwendungsgebiet ALGS bereits bewertet. Für PFIC konnten dagegen keine Validierungsstudien identifiziert werden. In den Studienunterlagen war zur Bestimmung einer klinischen Relevanzschwelle für den ItchRO(Obs) eine verblindete Analyse vor Datenbankschluss anhand der Daten der Studie MARCH-PFIC vorgesehen. Weitere Details oder Ergebnisse konnten den Unterlagen des pU nicht entnommen werden. Dementsprechend kann die Validität des Instruments bei Personen mit PFIC

nicht abschließend bewertet werden. Da die Bewertung der Pruritus-Schwere (Item 1) anhand einer eindimensionalen, einfachen Skala erfolgt, ist von einer Übertragbarkeit auf das vorliegende Anwendungsgebiet auszugehen, sodass die Ergebnisse in der Nutzenbewertung berücksichtigt werden. Es liegen keine Validierungsdaten für die weiteren Items des ItchRO vor (auch nicht für Personen mit ALGS), wie z. B. für die Frequenz des Juckens (Item 3 des ItchRO(Obs)). Da die Frequenz des Juckens zudem als Teilaspekt der Pruritus-Schwere (Item 1 – Rubbeln/Kratzen) erachtet wird, werden diese Ergebnisse nicht gesondert dargestellt. Auch für den 4-Wochen-Durchschnittswert existieren keine Validierungsdaten.

Hinsichtlich der klinischen Relevanz ist die prädefinierte Responder-Definition von  $\geq 1$  Punkt Verbesserung oder einer Ausprägung von  $\leq 1$  Punkt inhaltlich nachvollziehbar. Weiterhin entspricht eine Verbesserung um mindestens 1 Punkt mehr als 15 % der Skalenspannweite (0–4 Punkte). 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. Zudem besteht die Möglichkeit, dass aufgrund des gewählten Einschlusskriteriums von  $\geq 1,5$  Punkte im ItchRO zu Baseline, auch Personen mit Verbesserungen von weniger als 1 Punkt als Responder deklariert wurden. Für Werte  $\leq 1$  ist davon auszugehen, dass fast keine Belastung durch den Pruritus vorliegt. Eine weitere Responderanalyse bezieht sich auf alle Personen, die für mindestens 50 % der Tage in dem jeweils betrachteten Zeitraum einen morgendlichen Pruritus-Schweregrad von  $\leq 1$  Punkt erreicht haben. Eine Begründung und Validierung für den Schwellenwert von 50% liegt nicht vor.

### ***Klinische Kratzskala***

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

#### Operationalisierung

##### *Beschreibung*

Pruritus wurde in den Studien MARCH-PFIC und RISE 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 MARCH-PFIC kontinuierlich über die mittlere Veränderung von Baseline zum jeweiligen Erhebungszeitpunkt. Responderanalysen waren nicht vorgesehen.

Für die Studie RISE erfolgte die Auswertung rein deskriptiv für die Veränderung gegenüber Baseline.

##### *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 [19]. Der Publikation konnten jedoch keine Informationen zur Entwicklung oder formalen Validierung entnommen werden. Kritisch ist zu sehen, dass das Instrument innerhalb der 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.

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 für den überwiegenden Teil der Studienpopulation schwierig, sodass eine Einschätzung durch eine dritte Person für die gesamte Studienpopulation 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 Betreuer-bewertung verstärkt Hautverletzungen 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 RISE, 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 [8].

### ***Eindruck der Schwere (CIS/PIS)***

Der Endpunkt „Eindruck der Schwere“ wird in der Nutzenbewertung aufgrund einer Doppelerfassung des Juckreizes zum Endpunkt „ItchRO“ nicht berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Die Einschätzung zum Eindruck der Schwere des Juckreizes wird im Rahmen des „Caregiver Impression of Severity“ (CIS) durch die betreuende Person durchgeführt, während beim „Patient Impression of Severity“ (PIS) Patientinnen und Patienten im Alter von  $\geq 9$  Jahren selbst eine Einschätzung abgeben.

Beide Versionen (PIS und CIS) bestehen aus nur 1 Item. Der PIS fragt nach der patientenbewerteten Schwere des Pruritus in Bezug auf die vergangenen 7 Tage, mit den Antwortmöglichkeiten von „Ich verspürte gar keinen Juckreiz“ bis „Ich verspürte sehr starken Juckreiz“ auf einer 5-Punkte-Skala, wobei höhere Werte einen stärkeren Pruritus widerspiegeln. Der CIS hingegen fragt nach der durch die betreuende Person eingeschätzten Schwere des Pruritus innerhalb der vergangenen 7 Tage, wobei die Antwortmöglichkeiten von „Keine“ bis „Sehr stark“ auf einer 7-Punkte-Skala reichen, mit höherem Pruritus bei höheren Skalenwerten. Die Einschätzung der oder des Betreuenden kann dabei auf Äußerungen des Kindes oder auf mit dem Pruritus in Verbindung stehenden beobachtbaren Symptomen oder Verhaltensweisen beruhen.

Die Auswertung erfolgt über die mittlere Veränderung in CIS- und PIS-Werten zwischen Baseline und den Erhebungszeitpunkten. Responderanalysen waren nicht vorgesehen.

### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Die Schwere des Juckreizes wird bereits über den primären Endpunkt der Studie MARCH-PFIC „ItchRO“ valide erfasst. Die Fragestellungen sowie die Antwortoptionen sind fast deckungsgleich zu Item 1 des Patiententagebuchs „ItchRO“. Darüber hinaus liegen keine Responderanalysen für PIS/CIS vor. Insgesamt wird der Endpunkt nicht für die Nutzenbewertung berücksichtigt.

### Patientenrelevanz

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

### Validität

Da der Endpunkt aufgrund von Doppelerfassung nicht für die Nutzenbewertung berücksichtigt wird, erfolgt keine Bewertung der Validität.

### **EDQ – Juckreiz-Schweregrad-Score**

Der Endpunkt „EDQ“ wird in der Nutzenbewertung aufgrund von einer Doppelerfassung des Juckreizes zum Endpunkt „ItchRO“ nicht berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Das elektronische Tagebuch „Exploratory Diary Questionnaire“ (EDQ) wird analog zum ItchRO zweimal täglich ausgefüllt und erfasst die Schwere des Pruritus, Charakteristika des Kratzens sowie die Auswirkungen des Pruritus auf den Schlaf. Für Kinder < 9 Jahre wird das Instrument von der betreuenden Person ausgefüllt (EDQ(Obs)), Patientinnen und Patienten ≥ 9 Jahre füllen das Patienteninstrument aus (EDQ(Pt)). Die Studienteilnehmenden und die betreuenden Personen werden während des Screenings (Visite 0) in der Anwendung des elektronischen Tagebuchs geschult.

Basierend auf Beobachtungen oder Äußerungen des Kindes gegenüber der betreuenden Person fragt der morgendliche EDQ(Obs) danach, ob es dem Kind in der vergangenen Nacht schwerfiel, dem Kratzen oder Rubbeln zu widerstehen (Item 1), ob es durch das Kratzen zu blutenden Stellen kam (Item 2), ob es durch den Pruritus zu Durchschlafstörungen kam (Item 3) und wie schlimm der schwerste Juckreiz war (Item 4). Für die Items 1–3 konnte zwischen den Antwortoptionen 1 = „Nie“ bis 5 = „Fast immer“ gewählt werden, für Item 4 waren die Antwortoptionen 1 = „Kein Juckreiz“ bis 5 = „Sehr schwer“. Abends werden dieselben Fragen beantwortet, mit Ausnahme der Frage zu den Durchschlafstörungen (Item 3). Gemäß SAP war lediglich eine Auswertung des Items zum schwersten Juckreiz (Item 4) vorgesehen und des Items zu Durchschlafstörungen (siehe Endpunkt „Schlafstörungen mittels EDQ(Obs)“). Responderanalysen zum schwersten Juckreiz wurden nicht durchgeführt.

Der morgendliche patientenberichtete EDQ(Pt) besteht lediglich aus 1 Item und fragt nach der maximal erreichten Schwere des Juckreizes in der vergangenen Nacht (mit 1 = „Kein Juckreiz“ bis 5 = „Sehr schwer“). Abends kommt zu dieser Frage noch ein weiteres Item hinzu, welches auf die Häufigkeit des Pruritus abzielt (mit 1 = „Nie“ bis 5 = „Fast immer“). Gemäß Studienunterlagen war eine Auswertung des EDQ(Pt) nicht geplant.

### *Bewertung*

Die Operationalisierung ist nur teilweise nachvollziehbar. Eine Auswertung des EDQ(Pt) war trotz Datenerhebung nicht vorgesehen. Weiterhin ist unklar, weshalb der EDQ(Obs) analog zum ItchRO(Obs) nicht für alle Patientinnen und Patienten beantwortet wurde, sondern nur für jene unter 9 Jahre. Ergebnisse liegen somit nur für eine Teilpopulation vor. Da zudem die Schwere des Juckreizes bereits über den Endpunkt „ItchRO“ für alle Patientinnen und Patienten erfasst wurde, wird der Juckreiz mittels EDQ nicht für die Nutzenbewertung berücksichtigt.

### Patientenrelevanz

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

### Validität

Da der Endpunkt aufgrund von Doppelerfassung nicht für die Nutzenbewertung berücksichtigt wird, erfolgt keine Bewertung der Validität.

### **Schlafstörungen mittels EDQ(Obs)**

Der Endpunkt „Schlafstörungen mittels EDQ(Obs)“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Um mögliche Verbesserungen des Schlafs zu beurteilen, wird die schlafbezogene Frage aus dem EDQ(Obs) ausgewertet (siehe „EDQ – Juckreiz-Schweregrad-Score“). Beim EDQ(Obs) wird die Schlafstörung im Morgen-Tagebuch durch die Frage „Aufgrund von Juckreiz hatte mein Kind Schwierigkeiten durchzuschlafen“ auf einer 5-Punkte-Likert-Skala bewertet:

- 1 = Nie
- 2 = Selten
- 3 = Manchmal
- 4 = Oft
- 5 = Fast immer

Für die schlafbezogene Frage des EDQ(Obs) werden analog zum ItchRO 6-Wochen-Durchschnittswerte für die Wochen 1–6 und 4-Wochen-Durchschnittswerte für die verbleibenden Zeiträume ermittelt. Auch die Ermittlung des Baseline-Werts (4-Wochen-Durchschnittswert) erfolgt analog zum ItchRO.

Laut SAP sind folgende Auswertungen definiert:

- Die mittlere Veränderung des durchschnittlichen Werts zwischen Baseline und Woche 6, 10, 14, 18, 22 und 26 sowie der Durchschnitt der Wochen 15–26.
- Der Anteil der Personen (Responder) mit jeglicher Verbesserung im durchschnittlichen Wert gegenüber Baseline zu den Wochen 6, 10, 14, 18, 22 und 26.
- Der Anteil der Personen mit einem durchschnittlichen Wert  $\leq 2$  Punkte von Baseline bis Woche 26.

#### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar. Unklar ist, warum Schlafstörungen lediglich im betreuerberichteten Tagebuch EDQ(Obs) erhoben wurden und nicht auch im EDQ(Pt). Weiterhin ist unklar, weshalb der EDQ(Obs) analog zum ItchRO(Obs) nicht für alle Patientinnen und Patienten beantwortet wurde, sondern nur für jene unter 9 Jahre. Ergebnisse liegen somit nur für eine Teilpopulation vor.

### Patientenrelevanz

Pruritus kann zu Ein- und Durchschlafstörungen führen. Mit den Items des elektronischen Tagebuchs werden die Auswirkungen von Schlafmangel und Müdigkeit auf alltägliche Aktivitäten oder das Verhalten von Testpersonen nicht erfasst. Insofern kann die aus einer Tagesmüdigkeit resultierende Belastung der Patientinnen und Patienten nicht bewertet werden. Die Patientenrelevanz von Müdigkeit allein bzw. dem Schlafverhalten der Testpersonen (nächtliches Aufwachen) ist unklar.

### Validität

Zum EDQ wurden vom pU keine Validierungsstudien eingereicht und auch eine orientierende Recherche konnte keine Informationen zur Entwicklung oder formalen Validierung identifizieren. Der pU selbst beschreibt den EDQ als explorativ.

Da die Bewertung der Schlafstörungen anhand einer eindimensionalen, einfachen Skala erfolgt, werden die Ergebnisse in der Nutzenbewertung dennoch berücksichtigt. Allerdings können allein durch die Frage nach dem „Durchschlafen“ weitere Schlaf-assoziierte Probleme wie Einschlafschwierigkeiten oder eine generell verkürzte Schlafdauer nicht direkt erfasst werden.

Als Responder in den Auswertungen des pU sind zum einen alle Personen mit jeglicher Verbesserung der Durchschnittswerte im Vergleich zu Baseline definiert sowie diejenigen mit einem durchschnittlichen Wert  $\leq 2$  Punkte. Eine Begründung für die Ableitung der Responder-Definitionen liegt nicht vor. Eine Wertung jeglicher Verbesserung im durchschnittlichen Wert als Responder erscheint nicht adäquat. Ein Wert  $\leq 2$  entspricht den Antwortkategorien einer (1; Nie) nicht vorhandenen oder (2) seltenen Schlafstörung und erscheint als Responsekriterium prinzipiell plausibel. Allerdings liegen keine Angaben zum Anteil der Personen mit Werten  $\leq 2$  zu Baseline vor, die auch ohne Verbesserung als Responder in die Auswertung eingehen könnten.

Grundsätzlich wird die Selbsteinschätzung der betroffenen Person zu ihrem Krankheitszustand im Vergleich zu einer Fremdeinschätzung für die Nutzenbewertung bevorzugt. Der EDQ(Obs) wurde jedoch ausschließlich bei Kindern  $< 9$  Jahre eingesetzt, sodass im vorliegenden Fall 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.

### **sBA**

Der Endpunkt „sBA“ wird in der Nutzenbewertung aufgrund fehlender Patientenrelevanz nicht für die Nutzenbewertung berücksichtigt, aus Transparenzgründen jedoch im Anhang dargestellt.

### Operationalisierung

#### *Beschreibung*

Nüchternwerte der Serum-Gallensäure (sBA) werden in der Studie MARCH-PFIC bei jeder Studienvsiste im Rahmen einer Blutprobe als Cholestase-Biomarker durch das Studienpersonal erhoben, außer bei Visite 3.

Untersucht wird die mittlere Änderung des Gesamt-sBA-Spiegels zwischen Baseline und dem Durchschnitt der Wochen 18, 22 und 26 sowie der Anteil der sBA-Responder von Woche 18 bis Woche 26. Ein sBA-Responder ist definiert als eine Person mit durchschnittlichem sBA-Spiegel  $< 102 \mu\text{mol/l}$  (gilt nur, wenn der Baseline-sBA-Spiegel  $\geq 102 \mu\text{mol/l}$  war) ODER einer durchschnittlichen prozentualen Verringerung um mindestens 75 % gegenüber Baseline. Baseline ist definiert als der Durchschnitt der Daten, die zwischen Screening und Visite 1 erhoben wurden, oder als der letzte nicht fehlende Messwert vor Erhalt der ersten Dosis des Studienmedikaments, falls nur ein Wert verfügbar ist. Falls alle Werte fehlen, wird die Person als Non-Responder gewertet.

In der Studie RISE wird die Änderung des Nüchtern-sBA-Werts von Baseline zu Woche 13 untersucht. Baseline ist definiert als die letzten beiden verfügbaren Werte vor der ersten Gabe der Studienmedikation, oder als der letzte verfügbare Wert, falls nur ein Wert verfügbar ist.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

Beim sBA-Spiegel handelt es sich um einen Laborparameter, der, basierend auf einer Blutprobe, bei fast jeder Studienvsiste gemessen wird (außer bei Visite 3). Der pU beschreibt sBA als einen Cholestase-Biomarker und stellt die Ergebnisse in Modul 4 ergänzend beim Endpunkt „Pruritus“ dar. Als Laborparameter ist der Endpunkt nicht per se patientenrelevant. Es wurden keine Daten eingebracht, die den sBA-Spiegel als Surrogat für Pruritus qualifizieren. Da der sBA-Wert im vorliegenden Indikationsgebiet ein zentraler Parameter ist, erfolgt aus Transparenzgründen eine Darstellung im Anhang.

### Validität

Die Erhebung des Endpunkts wird als valide angesehen. Eine Begründung für die Festlegung der entsprechenden ResponderSchwellen führt der pU nicht an.

### **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 Weltgesundheitsorganisation (WHO) [20] zur Ableitung der z-Scores herangezogen und ab einem Alter von 24 Monaten die Wachstumskurven der „Centers for Disease Control and Prevention“ (CDC) [3].

In der Studie MARCH-PFIC wird die Körpergröße von geschultem Personal vor Ort mit einem kalibrierten Stadiometer (bei Kindern, die selbständig stehen können, i. d. R.  $\geq 2$  Jahre) bzw. einem geeichten Kopfbrett/Kopfteil (bei Kindern, die nicht selbständig stehen können, i. d. R.  $< 2$  Jahre) gemessen. Es erfolgen 2 unabhängige Messungen, die auf 0,1 cm genau aufgezeichnet werden. Eine 3. Messung erfolgt, wenn die Werte der 1. und 2. Messung um mehr als 0,5 cm voneinander abweichen. Für alle Messungen im Rahmen der Studie sollte dasselbe Stadiometer verwendet werden. Das Körpergewicht wird für Kinder, die selbständig stehen können (im Allgemeinen  $\geq 2$  Jahre), mit einer geeichten oder einer elektronischen Waage auf 0,1 kg genau ermittelt. Für Kinder  $< 2$  Jahre wird eine Säuglingswaage genutzt. Den Studienzentren werden über das Studienmanual standardisierte Instruktionen zur Erhebung von Körpergröße und Körpergewicht bereitgestellt. Untersucht wurde die mittlere Veränderung im Vergleich zu Baseline.

In der Studie RISE sind die Messungen ebenfalls dahingehend spezifiziert, dass sie nach standardisierten Vorgaben erfolgen sollen, die den Studienzentren zur Verfügung gestellt werden. Die Körpergröße (Länge) wird mit einem geeichten Kopfteil durch 2 unabhängige Messungen auf 0,1 cm genau gemessen (und durch eine 3. Messung, wenn die Werte um mehr als 0,5 cm abweichen). Das Körpergewicht wird mit einer geeichten Säuglingswaage durch 2 unabhängige Messungen ermittelt und auf 0,1 kg genau aufgezeichnet (und durch eine 3. Messung, wenn die Werte um mehr als 0,3 kg abweichen).

Die Auswertung erfolgte über die mittlere Veränderung des z-Scores für Körpergröße und Körpergewicht zu den jeweiligen Erhebungszeitpunkten gegenüber Baseline (für Studie MARCH-PFIC zusätzlich zu Woche 18, 22 und 26 kombiniert). Responderanalysen waren nicht vorgesehen.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

Wachstumsstörungen treten häufig bei Kindern mit PFIC auf [1,2]. 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

Die Erhebung der körperlichen Entwicklung erscheint valide.

Die Auswertung für die Studien MARCH-PFIC und RISE erfolgt über die mittlere Veränderung des Z-Scores für Körpergröße und Körpergewicht gegenüber dem Baseline-Wert. Vom pU wurde keine Evidenz zu einer Relevanzschwelle hinsichtlich der Veränderungen vorgelegt.

### **Leberassoziierte Ereignisse**

Der Endpunkt „Leberassoziierte Ereignisse“ wird in der Nutzenbewertung ergänzend dargestellt.

### Operationalisierung

#### *Beschreibung*

Leberassoziierte Ereignisse waren in der Studie MARCH-PFIC definiert als PEBD-Operationen (Partielle externe Gallengangsableitung), Aufnahme auf die Warteliste für Lebertransplantation, Leberdekompensation (hepatische Enzephalopathie, Varizenblutung, Aszites und spontane bakterielle Peritonitis), hepatozelluläres Karzinom (HCC) oder Tod. Spezifische Erhebungszeitpunkte sind in den Studienunterlagen nicht definiert. Ein Teil der definierten Ereignisse, wie die Aufnahme auf die Warteliste für eine Lebertransplantation oder PEBD-Operationen, führen jedoch direkt zu einem Abbruch der Studienteilnahme.

Die Auswertung erfolgt über die Zeit bis zu den leberassoziierten Ereignissen.

#### *Bewertung*

Es ist unklar, zu welchem Zeitpunkt und in welchem Rahmen die Erhebung des Endpunkts erfolgte.

### Patientenrelevanz

Bei „Leberassoziierte Ereignisse“ handelt es sich um einen zusammengesetzten Endpunkt. Während einige Ereignisse patientenrelevant erscheinen, wie PEBD-Operation, HCC oder Tod, ist die Patientenrelevanz einer Aufnahme auf die Warteliste für Lebertransplantation unklar.

### Validität

Es erscheint fraglich, ob mit der Operationalisierung alle existierenden leberassoziierten Ereignisse erfasst werden können. Der pU berücksichtigt den Endpunkt im Dossier nicht und es liegen keine Informationen oder Begründungen vor, inwiefern die gelisteten Ereignisse von vergleichbarem Schweregrad sind.

## **PedsQL-Fatigue**

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

### Operationalisierung

#### *Beschreibung*

Zusätzlich zum „Pediatric Quality of Life Inventory“ (PedsQL) wurde in der Studie MARCH-PFIC die „PedsQL – Multidimensional Fatigue Scale“ (PedsQL-Fatigue), Version 3.0, eingesetzt; ein 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: Patienten- und Elternberichte in je einer Version für 8- bis 12-Jährige und 13- bis 18-Jährige sowie Elternberichte für Kleinkinder in je einer Version für 2–4 Jahre und 5–7 Jahre. Sowohl die elternberichtete als auch die selbstberichtete Version der PedsQL-Fatigue kam damit nur in Subgruppen der Patientinnen und Patienten zum Einsatz. Für die Auswertung wurde laut SAP jedoch nur die elternberichtete Version berücksichtigt.

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 5-Punkte-Likert-Skala von 0 (= nie; entspricht der besten Funktion) bis 4 (= fast immer; entspricht der schlechtesten Funktion). Die Werte werden anschließend in eine Skala von 0 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 zu Baseline wurde dieselbe Version während der gesamten Dauer der Studie verwendet.

Konkrete Angaben zur geplanten Auswertung der PedsQL-Fatigue konnten den Unterlagen des pU nicht entnommen werden. Eine Auswertung erfolgte für den Gesamtwert der betreuerberichteten Version des PedsQL-Kernmoduls gegenüber Baseline (siehe unten). Responderanalysen waren nicht vorgesehen.

#### *Bewertung*

Der pU ordnet die PedsQL-Fatigue der Lebensqualität zu. Dem wird nicht gefolgt und stattdessen eine Zuordnung zur Morbidität vorgenommen. Es ist nicht nachvollziehbar, warum gemäß Studienunterlagen lediglich eine Auswertung der mittleren Veränderung des elternberichteten Gesamtscores erfolgte.

### 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.

Es ist unklar, weshalb ein Selbstbericht durch Kinder nicht bereits im Alter von 5 bis 7 Jahren erfolgte.

### 2.3.3 Lebensqualität

#### **PedsQL**

Der Endpunkt „PedsQL“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Zur Erfassung der gesundheitsbezogenen Lebensqualität in der Studie MARCH-PFIC wurde der generische 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 [9]. In Abhängigkeit vom Alter der Testperson wurden in der Studie MARCH-PFIC unterschiedliche Versionen des PedsQL eingesetzt: Patienten- und Elternberichte in je einer Version für 8- bis 12-Jährige und 13- bis 18-Jährige sowie Elternberichte für Kleinkinder in je einer Version für 1–12 Monate, 13–24 Monate, 2–4 Jahre und 5–7 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.

Der Gesamtscore sowie die Subscores für körperliche Gesundheit, psychosoziale Gesundheit werden separat für die Berichte der Eltern und der Teilnehmenden berechnet.

Die Beantwortung der Items erfolgt auf einer 5-Punkte-Likert-Skala von 0 (= nie; entspricht der besten Funktion) bis 4 (= fast immer; entspricht der schlechtesten Funktion). 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.

Gemäß SAP ist lediglich eine Auswertung der mittleren Veränderung des elternberichteten Gesamtscores gegenüber Baseline geplant. Responderanalysen wurden nicht durchgeführt.

##### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar. Es ist nicht nachvollziehbar, dass gemäß SAP im Studienbericht lediglich eine Auswertung der mittleren Veränderung des elternberichteten Gesamtscores erfolgte.

#### 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 in allen Versionen bewertet. Zusammenfassend wird der PedsQL als geeignet zur Erfassung der gesundheitsbezogenen Lebensqualität in den vorliegenden Altersgruppen bewertet.

Es ist unklar, weshalb ein Selbstbericht durch Kinder nicht bereits im Alter von 5 bis 7 Jahren erfolgte.

## 2.3.4 Sicherheit

### Unerwünschte Ereignisse

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

#### Operationalisierung

##### *Beschreibung*

Unerwünschte Ereignisse (UE) sind in der Studie MARCH-PFIC definiert als jegliche ungünstigen und ungewollten medizinischen Zwischenfälle/Ereignisse in zeitlichem Zusammenhang mit der Einnahme der Studienmedikation; unabhängig davon, ob die Prüfmedikation kausal im Zusammenhang mit dem UE steht.

Alle UE werden ab dem Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis zum Ende der Follow-up-Periode erhoben. Für Teilnehmende, die sich vor der Follow-up-Visite zu Woche 27 in die Extensionsstudie MRX-503 einschreiben, werden die UE bis zum Zeitpunkt der Unterzeichnung der Einwilligungserklärung von MRX-503 erfasst.

Personen, die die Studie vorzeitig beenden, werden 7 Tage nach der letzten Dosis des Studienmedikaments telefonisch zu einer letzten Sicherheitsbeobachtung befragt. Personen, die die Studie aus Sicherheitsgründen abbrechen, werden ebenfalls solange nachbeobachtet, wie es klinisch angezeigt ist oder bis keine weitere Verbesserung eines UE zu erwarten ist. Alle UE, die zum Zeitpunkt dieser Kontaktaufnahme noch nicht behoben sind, werden von der Prüfstelle verfolgt, bis eines der oben genannten Kriterien erreicht ist.

PFIC-Symptome sollten nicht als UE eingestuft werden, solange sie innerhalb der normalen täglichen Schwankungen oder der erwarteten Progression der Erkrankung liegen und Teil der in der Studie zu erfassenden Wirksamkeitsdaten sind. Allerdings sollte eine Krankheitsprogression, die zum Studienabbruch führt, als UE erfasst werden.

Als Treatment-Emergent Adverse Events (TEAE) sind alle UE definiert, die während oder nach der Einnahme der ersten Dosis des Prüfpräparats und spätestens 7 Tage nach der letzten Dosis auftreten (für Patientinnen und Patienten, die nicht an der Extensionsstudie teilnehmen) oder bis zur Visite in Woche 26 / dem Ende der Behandlung berichtet werden (für Teilnehmende an der Extensionsstudie). Dazu zählen auch bestehende Ereignisse, die sich mit der Einnahme der Studienmedikation in Intensität oder Häufigkeit verstärkten.

Die UE wurden gemäß „Medical Dictionary for Regulatory Activities“ (MedDRA), Version 22.1, nach Systemorganklasse (SOC) und Preferred Term (PT) kodiert. Laut Dossier erfolgte die Einteilung in Schweregrade gemäß „Common Terminology Criteria for Adverse Events“ (CTCAE), Version 4.0; im Studienprotokoll ist jedoch eine Schweregradeinteilung anhand folgender Kategorien beschrieben:

- **mild:** Ein UE, das für gewöhnlich vorübergehend ist und nur minimale Behandlung oder therapeutische Intervention erfordert. Das Ereignis beeinträchtigt im Allgemeinen nicht die üblichen Aktivitäten des täglichen Lebens.
- **moderat:** Ein UE, das i. d. R. durch spezifische therapeutische Maßnahmen gelindert werden kann. Das Ereignis beeinträchtigt die üblichen Aktivitäten des täglichen Lebens und verursacht Unbehagen, stellt aber keine signifikante oder dauerhafte Gefahr für die/den Studienteilnehmende/n dar.
- **schwer:** Ein UE, das die üblichen Aktivitäten des täglichen Lebens unterbricht oder den klinischen Zustand erheblich beeinflusst oder intensive therapeutische Maßnahmen erfordern kann.

Als schwerwiegende unerwünschte Ereignisse (SUE) wurden UE definiert, die

- zum Tode führten,
- direkt lebensgefährdend waren,
- zur Hospitalisierung bzw. Verlängerung einer Hospitalisierung führten (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.

In den Studienunterlagen werden folgende Ereignisse von klinischem Interesse definiert, die im Dossier hingegen als „UE von besonderem Interesse“ bezeichnet werden:

- Diarrhö (definiert als PT und „Lowest Level Term“ (LLT) „Diarrhö“, „Intermittierende Diarrhö“, „Erhöhte Stuhlentleerung“ und „Gastroenteritis“).
- Ereignisse im Zusammenhang mit erhöhtem Bilirubin (definiert als PT und LLT „Bilirubin im Blut erhöht“, „Bilirubin im Blut anomal“ und „Hyperbilirubinämie“).
- Ereignisse im Zusammenhang mit erhöhten Transaminasen (PT und LLT „Erhöhte Leberenzyme“, „Transaminasen anomal“, „Transaminasen erhöht“, „Alanin-Aminotransferase erhöht“, „Aspartat-Aminotransferase erhöht“, „Aspartat-Aminotransferase anomal“ und „Alanin-Aminotransferase anomal“).
- Ereignisse im Zusammenhang mit Mängeln der fettlöslichen Vitamine (viele im Anhang des SAP gelistete, SOC-übergreifende PT).

In der Studie RISE 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 abnormale Laborergebnisse oder Anzeichen und Symptome gewertet, die mit der Grunderkrankung im Zusammenhang stehen.

Die Klassifikation von SUE erfolgte weitestgehend analog zur Studie MARCH-PFIC. 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.

Zudem erfolgte die Beurteilung des Schweregrads in der Studie RISE laut Studienprotokoll und SAP anhand CTCAE-Version 5.0, laut Dossier anhand Version 4.0.

### *Bewertung*

Die Operationalisierung in den Studien MARCH-PFIC und RISE ist weitestgehend nachvollziehbar. Im Dossier gibt der pU an, schwere UE nach CTCAE zu klassifizieren. In den Studienunterlagen von MARCH-PFIC ist dagegen eine studienindividuelle Schweregradeinteilung bestimmt, die sich an der Beeinträchtigung täglicher Aktivitäten orientiert. Im vorliegenden Fall wird die alternative Einteilung als sachgerecht beurteilt.

Erwartbare PFIC-Symptome innerhalb der täglichen Schwankungen sollten nicht als UE eingestuft werden. Eine Auflistung dieser PFIC-Symptome konnte nicht identifiziert werden. Darüber hinaus konnten keine Angaben identifiziert werden, ob Ereignisse der Grunderkrankung für die Auswertung berücksichtigt wurden.

### 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, solange sie innerhalb der normalen täglichen Schwankungen oder der erwarteten Progression der Erkrankung liegen und Teil der in der Studie zu erfassenden Wirksamkeitsdaten sind, nicht als UE erfasst werden. Die zugrundeliegenden Symptome sind nicht definiert. Es bleibt unklar, ob mit dieser Einschränkung ausreichend zwischen Grunderkrankung und Nebenwirkungen unterschieden werden kann.

Patientinnen und Patienten werden 7 Tage nach der letzten Dosis des Studienmedikaments telefonisch zu einer letzten Sicherheitsbeobachtung befragt. Eine Weiterverfolgung nach Studienabbruch über diese 7 Tage erfolgt nur bei bereits bestehenden UE. Es ist unklar, ob 7 Tage ausreichend sind, um alle UE zu erfassen, die potentiell mit einem zeitlichen Abstand zur letzten Gabe der Studienmedikation auftreten.

Gemäß SAP wurden explizit keine UE von besonderem Interesse bestimmt, sondern lediglich Ereignisse von „klinischem Interesse“, bestehend aus Diarrhö, Mangel an fettlöslichen Vitaminen und Ereignissen im Zusammenhang mit erhöhten Transaminasen und Bilirubin. Dabei geht der pU davon aus, dass Abweichungen bei fettlöslichen Vitaminen bzw. Transaminasen und Bilirubin mit der Grunderkrankung PFIC in Verbindung stehen und Maralixibat die gastrointestinalen Störungen verursacht. Bei der Interpretation der UE von besonderem Interesse ist zu beachten, dass zumindest teilweise Ereignisse der Grunderkrankung erfasst werden.

### 2.3.5 Erhebungszeitpunkte

Eine Übersicht der Erhebungszeitpunkte der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte findet sich für die Studie MARCH-PFIC in Tabelle 8 und für die Studie RISE in Tabelle 9.

*Tabelle 8: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie MARCH-PFIC*

Endpunkt	Studiensite (Woche)	Screening	Dosis- eskalation		Erhaltungsphase						Follow-up
	V0	V1 <sup>1)</sup>	V2	V3 <sup>2)</sup>	V4	V5	V6	V7	V8	EOT/ET <sup>3)</sup>	
	W-6	W0	W2	W4	W6	W10	W14	W18	W22	W26	W27
Todesfälle <sup>4)</sup>		x	x	x	x	x	x	x	x	x	x
ItchRO	Zweimal täglich ausgefüllt.										
EDQ <sup>5)</sup>	Zweimal täglich ausgefüllt.										
Klinische Kratzskala <sup>6)</sup>	x	x	x	x	x	x	x	x	x	x	
sBA	x	x	x		x	x	x	x	x	x	
Körperliche Entwicklung	x	x	x	x	x	x	x	x	x	x	
PedsQL		x		x	x	x	x	x	x	x	
Unerwünschte Ereignisse <sup>7)</sup>		x	x	x	x	x	x	x	x	x	x

<sup>1)</sup> Baseline-Erhebung.

<sup>2)</sup> Die Dosisescalation dauerte 4–6 Wochen. Es ist daher möglich, dass sich zu Visite 3 in Woche 4 einige der Teilnehmenden noch in der Dosisescalationsphase befanden, andere hingegen schon ihre stabile Dosis erhielten.

<sup>3)</sup> Bei vorzeitigem, dauerhaftem Abbruch der Studienmedikation wurde die/der Teilnehmende aus der Studie ausgeschlossen und es wurden die für Woche 26 (EOT/ET) aufgelisteten Untersuchungen durchgeführt. Die ET-Untersuchungen sollten bei dem Besuch durchlaufen werden, bei dem auch die Studienmedikation abgesetzt wurde, oder es sollte so bald wie möglich ein zusätzlicher ET-Besuch angesetzt werden, falls die Studienmedikation zwischen den Besuchen abgesetzt wurde. Teilnehmende, die vorzeitig aus der Studie ausschieden, erhielten den für Woche 27 angesetzten Sicherheits-Follow-up-Anruf 7 ± 2 Tage nach der letzten Dosis der Studienmedikation.

<sup>4)</sup> Im Rahmen der Sicherheit erhoben.

<sup>5)</sup> Da der Endpunkt „Schlafstörungen“ anhand des morgendlichen EDQ erhoben wird, ist dieser nicht separat in der Tabelle aufgeführt.

<sup>6)</sup> Endpunkt wird ergänzend dargestellt.

<sup>7)</sup> In den Wochen zwischen den Studiensiten (Woche 1, 3, 5, 8, 12, 16, 20 und 24) gab es zusätzliche Kontakte via Telefon oder E-Mail mit den Teilnehmenden zur Erfassung von UE sowie von Vor- und Begleittherapien.

Abkürzungen: EDQ: Exploratory Diary Questionnaire; EOT: End of Treatment; ET: Early Termination; ItchRO: Itch Reported Outcome; PedsQL: Pediatric Quality of Life Inventory; sBA: Serum-Gallensäure; UE: Unerwünschtes Ereignis; V: Visite; W: Woche.

*Tabelle 9: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie RISE*

Endpunkt \ Studiervisite (Woche)	Baseline	Dosis- eskalation <sup>1</sup>		Erhaltungs- phase	Ende Kern- studie <sup>2</sup>	LTE <sup>3</sup>		ET <sup>4</sup>	Follow- up
		W3	W6			W10	W13		
Todesfälle <sup>5</sup>	x	x	x	x	x	x	x	x	x
ItchRO(Obs) <sup>6</sup>	x	x	x	x	x	x		x	
Klinische Kratzskala <sup>7</sup>	x		x		x	x	x	x	
sBA	x	x		x	x	x	x	x	
Körperliche Entwicklung	x	x	x	x	x	x	x	x	
Unerwünschte Ereignisse	x	x	x	x	x	x	x	x	x

<sup>1</sup>) In der Dosisescalationsphase 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 7–13).

<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.

<sup>6</sup>) Nach Baseline jeweils 7 Tage vor der entsprechenden Visite von den betreuenden Personen auszufüllen.

<sup>7</sup>) Endpunkt wird ergänzend dargestellt.

Abkürzungen: ET: Early Termination; LTE: Langfristige Behandlungsverlängerung; ItchRO(Obs): Itch Reported Outcome (Observer); sBA: Serum-Gallensäure; W: Woche.

## 2.4 Statistische Methoden

Die statistischen Methoden basieren auf dem SAP der Studie MARCH-PFIC, der in der finalen Version (1.0) auf den 21.09.2022 datiert ist, sowie auf dem SAP der Studie RISE, dessen finale Version (2.0) auf den 20.03.2024 datiert ist.

### Analysepopulationen

- Intention to Treat (ITT)-Population: Alle randomisierten Personen.
- Sicherheitspopulation: Alle eingeschlossenen Personen, die mindestens eine Dosis der Studienmedikation erhalten haben.

Die Studie MARCH-PFIC definiert laut SAP zudem folgende Populationen für die Auswertung:

- Primäre-Population:
  - Umfasst alle Teilnehmenden mit nt-PFIC2.
  - Schließt alle anderen PFIC-Subtypen aus sowie Teilnehmende mit Heterozygotie, mit niedrigen oder schwankenden sBA-Werten, mit vorheriger Operation zur Behandlung von PFIC, mit t-PFIC2 sowie Teilnehmende, bei denen keine etablierte mit PFIC-assoziierte genetische Variante identifiziert werden kann.
- PFIC-Population:
  - Umfasst alle Teilnehmenden mit PFIC1, nt-PFIC2, PFIC3, PFIC4 und PFIC6, d. h. alle bekannten etablierten Genotypen außer t-PFIC2.
  - Schließt Teilnehmende mit Heterozygotie, mit niedrigen oder schwankenden sBA-Werten oder mit vorheriger Operation zur PFIC-Behandlung aus sowie Teilnehmende, bei denen keine etablierte mit PFIC-assoziierte genetische Variante identifiziert werden kann.
- Gesamtpopulation: Umfasst alle Teilnehmenden der Studie MARCH-PFIC.

### Datenschnitte

Die berücksichtigten Analysen der Studie MARCH-PFIC beziehen sich auf den finalen Datenschnitt vom 14.10.2022.

Für die Studie RISE war eine Interimsanalyse geplant, wenn mindestens 6 Kinder die 13-wöchige Kernstudie durchlaufen haben. Zum vorliegenden Datenschnitt vom 23.06.2022 wurde diese lediglich von 4 Kindern mit PFIC abgeschlossen. Ergebnisse für eine Interimsanalyse liegen vor, jedoch nur für 4 Personen mit PFIC. Im European Public Assessment Report (EPAR) sind Ergebnisse eines weiteren Interim-Datenschnitt der Studie RISE vom 28.06.2023 enthalten, die vom pU jedoch nicht vorgelegt wurden. Eine finale Analyse existiert noch nicht.

### Präspezifizierte Subgruppenanalysen

In der Studie MARCH-PFIC waren keine Subgruppenanalysen für die im Rahmen der Nutzenbewertung berücksichtigte PFIC- und Gesamtpopulation präspezifiziert.

Dem Protokoll der Studie RISE konnten keine Subgruppenanalysen entnommen werden.

### Präspezifizierung geplanter und durchgeführter Analysen

#### Kontinuierliche Endpunkte

In der Studie MARCH-PFIC wurden kontinuierliche, im Zeitverlauf erhobene, Wirksamkeitsvariablen mittels „Mixed Model for Repeated Measures“ MMRM analysiert. Für die PFIC-Population wird auch der PFIC-Typ als Kovariate in der MMRM-Analyse verwendet. Aufgrund der geringen Stichprobengrößen werden für diese Analyse PFIC4, PFIC5 und PFIC6 zusammengefasst. Innerhalb der PFIC-Population werden 4 Ebenen von PFIC-Typen als Kovariate einbezogen: nt-PFIC2, PFIC1, PFIC3 und andere (PFIC4, PFIC5 und PFIC6 als Gruppe). Keine weiteren Kovariaten werden in die Analysen der Wirksamkeitsendpunkte einbezogen.

### Binäre Endpunkte

Für die Responderanalysen bei Wirksamkeitsendpunkten berechnet der pU zusätzlich zu den präspezifizierten p-Werten mittels Barnard's exaktem Test post hoc einen p-Wert mittels Cochran-Mantel-Haenszel (CMH)-Test.

### ItchRO

Zusätzlich zu den in Kapitel 2.3.2 aufgelisteten Auswertungen für den ItchRO präsentiert der pU im Nutzendossier 2 weitere Post-hoc-Operationalisierungen:

- Anteil der Personen mit einem mittleren morgendlichen ItchRO(Obs)-Schweregrad-Score  $\leq 1$  zu Woche 26.
- Anteil der Tage mit einem Wert von  $\leq$  Punkt im ItchRO(Obs) in der Erhebung am Morgen für die kumulierte Anzahl der Tage, in denen sich alle Personen im Studienzeitraum befanden.

Die Auswertungen in der Studie RISE erfolgten ausschließlich deskriptiv.

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

Beim primären Endpunkt „Durchschnittlicher Schweregrad des morgendlichen Juckreizes (ItchRO(Obs))“ der Studie MARCH-PFIC wurden zum Umgang mit fehlenden Daten neben der primären Analyse mittels MMRM zusätzlich noch Sensitivitätsanalysen mittels multipler Imputation unter Verwendung des „Standard-Missing-at-Random (MAR)“-Ansatzes und mittels multipler Imputation unter Verwendung des Tipping-Point-Ansatzes bestimmt.

Fehlende Werte bei den Tagebüchern ItchRO und EDQ waren wie folgt definiert: Für wöchentliche Morgen-/Abend-/Höchstwerte der Schwere und/oder der Häufigkeit (ItchRO/EDQ) gilt der Durchschnittswert für diesen wöchentlichen Zeitraum als fehlend, wenn mehr als die Hälfte (d. h.  $> 3$  von 7) der Tageswerte nicht verfügbar waren. Bei ItchRO-/EDQ-Werten für 4- und 6-wöchige Auswertungszeiträumen gilt ebenfalls der jeweilige Wert für den Zeitraum als fehlend, wenn mehr als die Hälfte ( $\geq 2$  von 4 bzw.  $\geq 3$  von 6) der wöchentlichen Werte fehlten.

Wenn bei der Ableitung des höchsten täglichen Schweregrads- und/oder Häufigkeitswerts entweder der Morgen- oder der Abendbericht nicht ausgefüllt wurde, wurde der jeweils verfügbare Wert als höchster Tageswert verwendet. Wenn sowohl der Morgen- als auch der Abendbericht nicht ausgefüllt wurden, galt der höchste Tageswert für diesen Tag als fehlend.

### Fehlende Werte bei Responderanalysen

Im Rahmen von Responderanalysen werden fehlende Werte prinzipiell als Non-Responder behandelt. Für die Definition von fehlenden Werten siehe auch Kapitel 2.3.2.

Für die Auswertung der ItchRO-Responder werden Personen von Woche 15 bis Woche 26 als Non-Responder definiert, wenn der 4-Wochen-Durchschnittswert der Baseline fehlt ODER alle drei 4-Wochen-Durchschnittswerte fehlten (d. h. Wochen 15–18, 19–22 und 23–26).

Für die Anzahl der Tage, an denen der ItchRO(Obs)-Schweregrad des morgendlichen Juckreizes in den letzten 84 Tagen (d. h. Studientag 99 bis einschließlich 182, bezogen auf den Tag der ersten Dosis des Studienmedikaments)  $\leq 1$  ist, wird bei Tagen mit fehlenden Daten das Kriterium „ $\leq 1$ “ für diesen Tag als nicht erfüllt erachtet.

Bei der Berechnung von Anzahl und Anteil der Teilnehmenden mit einem morgendlichen ItchRO(Obs)-Schweregrad von  $\leq 1$  an mehr als 50 % der Tage (innerhalb der vorgegebenen Zeiträume), wird das Kriterium „ $\leq 1$ “ an Tagen mit fehlenden Daten als nicht erfüllt erachtet.

Für den Anteil der Teilnehmenden, bei denen sich die EDQ-Schlafwerte gegenüber dem Ausgangswert verbessert haben, wurden Teilnehmende als nicht verbessert (Non-Responder) definiert, wenn entweder deren 4-Wochen-Durchschnittswert zu Baseline fehlt ODER der 4- bzw. 6-Wochen-

Durchschnitt zu den Post-Baseline-Analysezeitpunkten. Für den Anteil der EDQ-Schlaf-Scores  $\leq 2$  werden fehlende Scores als Non-Responder behandelt (d. h.  $> 2$ ).

In der Studie RISE erfolgten keine relevanten Imputationen von fehlenden Werten.

### **Einschätzung der statistischen Auswertungen**

Eine Auswertung der Wirksamkeitsendpunkte in der Studie MARCH-PFIC für die Gesamtpopulation war nicht geplant und wurde auch post hoc nicht durchgeführt. Somit liegen keine Wirksamkeitsdaten in der für die Nutzenbewertung relevanten Gesamtpopulation vor. Aus diesem Grund werden ergänzend die Ergebnisse der PFIC-Population der Studie MARCH-PFIC dargestellt.

#### ItchRO

Im Nutzendossier des pU wurden die vielfältigen, prädefinierten Operationalisierungen des Endpunkts „ItchRO“ um zwei weitere post hoc durchgeführte Auswertungen ergänzt. Die Operationalisierung „Anteil der Personen mit einem mittleren morgendlichen ItchRO(Obs)-Schweregrad-Score  $\leq 1$  zu Woche 26“ bzw. deren Responder-Definition ist bereits Bestandteil der Responderanalyse „Anteil der ItchRO(Obs)-Responder zwischen Woche 15 und Woche 26“. A priori definierte Analysen, die sich auf das Erreichen eines Werts  $\leq 1$  Punkt im ItchRO(Obs) beziehen, liegen mit der „Anzahl Tage mit morgendlichem ItchRO(Obs)-Schweregrad-Score  $\leq 1$  über die letzten 84 Studientage“ und dem „Morgendlichen ItchRO(Obs)-Schweregrad-Score  $\leq 1$  für  $> 50\%$  des Zeitraums“ bereits vor. Da für die Post-hoc-Auswertung zudem eine ergebnisgesteuerte Berichterstattung nicht auszuschließen ist, werden Ergebnisse in der Nutzenbewertung nicht berücksichtigt.

Für die Nutzenbewertung werden weiterhin primär die Responderanalysen zum ItchRO als relevant erachtet. Zwar sind die beiden Responderkriterien „Verringerung um mindestens 1 Punkt im Vergleich zu Baseline“ und „Erreichen eines durchschnittlichen Schweregrads von  $\leq 1$ “ nicht formal validiert (siehe Kapitel 2.3.2). Eine Verbesserung um mindestens 1 Punkt entspricht jedoch mehr als 15 % der Skalenspannweite (0–4 Punkte). Auch das Erreichen eines Pruritus-Werts von  $\leq 1$  wird in der vorliegenden Nutzenbewertung jedoch als relevant erachtet, da dies keinem oder nur einem sehr geringen Pruritus entspricht. Allerdings liegen die Responderanalysen nur für den morgendlichen ItchRO(Obs)-Wert und nicht für den abendlichen bzw. den höchsten täglichen Wert vor, sodass ergänzend zu den Responderanalysen auch die stetigen Auswertungen über Mittelwertdifferenzen im Vergleich zu Baseline in der Nutzenbewertung für die morgendlichen, abendlichen und höchsten täglichen Werte berichtet werden.

Ergebnisse zur Häufigkeit 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 und zudem keine Responderanalysen für die Häufigkeit des Kratzens vorliegen.

#### Dichotome Auswertungen

Für die Responderanalysen bei Wirksamkeitsendpunkten stellt der pU neben den p-Werten mittels Barnard's exaktem Test post hoc zusätzlich p-Werte mittels CMH-Test dar. Im Folgenden werden für die Wirksamkeitsanalysen ausschließlich die p-Werte des Barnard's exaktem Test verwendet, da diese präspezifiziert waren und zudem keine Diskrepanzen bei den Ergebnissen ersichtlich sind. Weiterhin ist im Fall des CMH-Tests unklar, welche Strata verwendet wurden.

Bei den Auswertungen zur Sicherheit liegt ausschließlich ein post hoc berechneter p-Wert anhand des CMH-Tests vor. Nähere Angaben zum Test bzw. den verwendeten Strata konnten den Unterlagen des pU nicht entnommen werden.

Die Auswertungen in der Studie RISE erfolgten ausschließlich deskriptiv.

## 2.5 Verzerrungspotential auf Studien- und Endpunktebene

Die Studie MARCH-PFIC liefert randomisierte, kontrollierte Daten über einen Zeitraum von insgesamt 26 Wochen. Eine Übersicht über das Verzerrungspotential auf Studienebene erfolgt separat für beide Populationen.

Bei der Studie RISE handelt es sich um eine einarmige Studie, weshalb von einem hohen Verzerrungspotential auf Studien- und Endpunktebene ausgegangen wird.

Tabelle 10: Verzerrungspotential der Studie MARCH-PFIC

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
<b>MARCH-PFIC</b>							
PFIC-Population	Unklar <sup>1)</sup>	Ja	Ja	Ja	Nein	Ja <sup>2)</sup>	Unklar
Gesamtpopulation	Ja	Ja	Ja	Ja	Nein	Ja <sup>2)</sup>	Niedrig

<sup>1)</sup> Es wurde in 2 separate Kohorten randomisiert, in die primäre Kohorte und die ergänzende Kohorte (siehe Tabelle 2). Aus diesen beiden Kohorten wurde im Anschluss die PFIC-Population zur Analyse gebildet. Die Randomisierung erfolgte jedoch nicht für die Analysepopulation 1:1 auf die Behandlungsgruppen für beide Kohorten. In der ergänzenden Kohorte erfolgte die Randomisierung zudem stratifiziert nach PFIC1, PFIC3 und allen weiteren Studienteilnehmenden der Kohorte. Da sich die PFIC-Population aus den beiden Randomisierungspopulationen zusammensetzt, wobei nur Teile der ergänzenden Randomisierungspopulation für die PFIC-Population entnommen wurden, ist unklar, inwieweit die Randomisierung aufrechterhalten wird. Die Studiencharakteristika erscheinen größtenteils ähnlich zwischen den Behandlungsgruppen für beide Analysepopulationen.

<sup>2)</sup> Es liegen keine Angaben zur Beobachtungsdauer und zur Expositionsdauer in der Placebo-Gruppe vor. Aus Angaben zur Datenvollständigkeit des ItchRO(Obs) geht eine vergleichbare Beobachtungsdauer in den beiden Studienarmen hervor. Die Studienabbruchraten waren vergleichbar sowie gering zwischen den Studienarmen. Auf Studienebene ergibt sich daraus kein verzerrender Aspekt.

Abkürzungen: ItchRO(Obs): Itch Reported Outcome (Observer Version); PFIC: Progressive familiäre intrahepatische Cholestase.

Das Verzerrungspotential der Gesamtpopulation von Studie MARCH-PFIC auf Studienebene wird als niedrig eingeschätzt, für die PFIC-Population wird von einem unklaren Verzerrungspotential ausgegangen. Das Verzerrungspotential der relevanten Endpunkte für die PFIC-Population wird in Tabelle 11 dargestellt, für die Gesamtpopulation liegen nur deskriptive Ergebnisse zu UE vor.

**Tabelle 11: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie MARCH-PFIC; PFIC-Population<sup>1)</sup>**

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Todesfälle	Ja	Ja	Nein	Nein	Niedrig <sup>2)</sup>
Pruritus (ItchRO(Obs))	Ja	Unklar <sup>3)</sup>	Nein	Nein	Unklar <sup>4)</sup>
Körperliche Entwicklung	Ja	Ja	Nein	Ja <sup>5)</sup>	Hoch
Unerwünschte Ereignisse	Ja	Ja	Nein	Ja <sup>6)</sup>	Unklar <sup>4)</sup>

<sup>1)</sup> Das Verzerrungspotential der Endpunkte bezieht sich auf die PFIC-Population. Für die Gesamtpopulation liegen ausgenommen von den UE keine Wirksamkeitsergebnisse vor.

<sup>2)</sup> In der Gesamtpopulation sind ebenfalls keine Todesfälle aufgetreten.

<sup>3)</sup> Rücklaufquote nicht berichtet. Aus allgemeinen Angaben zum ItchRO-Tagebuch lässt sich eine ausreichende Datenvollständigkeit für den ItchRO(Obs) im gesamten Erhebungszeitraum ableiten. Für die einzelnen Analysen des ItchRO können jedoch keine konkreten Rücklaufquoten abgeleitet werden.

<sup>4)</sup> Verzerrungspotential auf Studienebene unklar.

<sup>5)</sup> Es ist zu berücksichtigen, dass die Kinder der Maralixibat-Gruppe zu Baseline beim Körpergewicht (z-Score) mit -1,75 stärker von der Norm abwichen als im Vergleich zu Placebo mit -1,28.

<sup>6)</sup> Es liegen keine Angaben zur Beobachtungsdauer und zur Expositionsdauer in der Placebo-Gruppe vor. Aus Angaben zur Datenvollständigkeit des ItchRO geht eine vergleichbare Beobachtungsdauer in den beiden Studienarmen hervor. Die Studienabbruchraten waren vergleichbar sowie gering zwischen den Studienarmen. Für den Endpunkt „Unerwünschte Ereignisse“ ergibt sich daraus kein verzerrender Aspekt.

Abkürzungen: ItchRO(Obs): Itch Reported Outcome (Observer Version); ITT: Intention to Treat; PFIC: Progressive familiäre intrahepatische Cholestase; UE: Unerwünschtes Ereignis.

Das Verzerrungspotential für die Endpunkte „ItchRO(Pt)“, „Schlafstörungen mittels EDQ(Obs)“ und „PedsQL“ ist nicht bewertbar, da aufgrund fehlender Rücklaufquoten keine Effektschätzer berichtet werden. Die Erhebung erfolgte jeweils nur in altersspezifischen Subpopulationen der PFIC-Population. Die weiteren Verzerrungsaspekte sind jedoch vergleichbar mit dem ItchRO(Obs).

### 3 Ergebnisse der eingeschlossenen Studien

#### 3.1 Studiencharakteristika und Studienmedikation

Die Ergebnisse der Studie MARCH-PFIC basieren auf dem finalen Datenschnitt vom 14.10.2022. Für die Studie RISE wird der Interim-Datenschnitt vom 23.06.2022 berücksichtigt.

In die Studie MARCH-PFIC wurden 93 Patientinnen und Patienten eingeschlossen und behandelt. Insgesamt befanden sich in der Studie 5 Geschwisterpaare, also insgesamt 10 Geschwisterkinder, wobei beide Kinder eines Geschwisterpaares verblindet in dieselbe Behandlungsgruppe randomisiert wurden. Der Anteil der Studienabbrüche war in beiden Behandlungsgruppen gering. Angaben zur Beobachtungsdauer in der Studie MARCH-PFIC fehlen. Hinsichtlich der Behandlungsdauer in der Studie MARCH-PFIC liegen Ergebnisse nur für den Maralixibat-Arm vor, die im Median sowohl für die PFIC-Population als auch für die Gesamtpopulation 183 Tage beträgt. Bezogen auf die Gesamtpopulation erhielten 95,7 % der Kinder die angestrebte Maximaldosis. Der Durchschnitt und der Median zu den täglichen Maralixibat-Dosen lassen Abweichungen von der täglichen Maximaldosis von 1.200 µg/kg erkennen. 74 Patientinnen und Patienten (79,6 %) gingen im Anschluss in die Extensionsstudie MRX-503 über.

Tabelle 12: Allgemeine Angaben; Studie MARCH-PFIC (finaler Datenschnitt)

Studie MARCH-PFIC Allgemeine Angaben	PFIC-Population		Gesamtpopulation	
	MRX	PBO	MRX	PBO
ITT-Population <sup>1)2)</sup> , n (%)	33 (100)	31 (100)	47 (100)	46 (100)
Sicherheitspopulation <sup>1)</sup> , n (%)	33 (100)	31 (100)	47 (100)	46 (100)
Anzahl der Geschwisterkinder, n (%)	4 (12,1)	4 (12,9)	4 (8,5)	6 (13,0)
Abbruch Einnahme der Studienmedikation, n (%)	1 (3,0)	3 (9,7)	3 (6,4)	4 (8,7)
Aufgrund von:				
UE	0	0	1 (2,1)	0
Krankheitsprogression	0	1 (3,2)	0	1 (2,2)
Entzug der Einverständniserklärung	1 (3,0)	2 (6,5)	1 (2,1)	3 (6,5)
Lebertransplantation	0	0	1 (2,1)	0
Übergang in die Extensionsstudie MRX-503	k. A.	k. A.	36 (76,6)	38 (82,6)
Mittlere Behandlungsdauer, Tage (SD) <sup>3)</sup>	184,4 (18,8)	k. A.	177,1 (36,6)	k. A.
Mediane Behandlungsdauer, Tage (min; max)	183 (108; 256)	k. A.	183 (12; 256)	k. A.
<i>Durchschnittliche Dosis (µg/kg/Tag)</i>				
MW (SD)	1.005,8 (146,5)	k. A.	962,4 (210,1)	k. A.
Median	1.041,8	k. A.	1.029,5	k. A.
(min; max)	(431,3; 1.126,6)	k. A.	(212,5; 1.126,6)	k. A.
Erhalt der Maximaldosis von 600 µg/kg BID, n (%)	k. A.	k. A.	45 (95,7)	k. A.
Mediane Beobachtungsdauer	k. A.	k. A.	k. A.	k. A.

<sup>1)</sup> Definition siehe Kapitel 2.4.

<sup>2)</sup> Eine Person wurde zur Baseline-Visite fälschlicherweise randomisiert. Diese Person erhielt kein Studienmedikament und brach die Studie, aufgrund eines „Administrationsfehlers“, ab. Einen Monat später wurde die Person erneut randomisiert; die Person wird jedoch nur einmal gezählt und taucht nicht bei den deskriptiven Ergebnissen zu Studienabbrüchen auf.

<sup>3)</sup> Behandlungsdauer = Datum der letzten Dosis des Studienmedikaments - Datum der ersten Dosis des Studienmedikaments + 1 Tag. Falls das Datum der letzten Dosis nicht bekannt war, wurde stattdessen das Datum des letzten bekannten Kontakts zur Berechnung der Behandlungsdauer verwendet.

Abkürzungen: BID: zweimal täglich; ITT: Intention to Treat; k. A.: keine Angabe; MRX: Maralixibat; MW: Mittelwert; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; SD: Standardabweichung; UE: Unerwünschtes Ereignis.

In der Studie RISE wurden alle 4 gescreenten Kleinkinder auch in die Studie eingeschlossen und mit Maralixibat behandelt. Zum Zeitpunkt des Interim-Datenschnitts haben alle Kinder die 12-wöchige Kernstudie beendet und sind in die langfristige Behandlungsphase ab Woche 13 übergegangen. Die mediane Behandlungsdauer lag zum Zeitpunkt des Datenschnitts bei 114,5 Tagen. Daten zur medianen Beobachtungsdauer legt der pU nicht vor. Angaben zur durchschnittlichen Tagesdosis konnten den Studienunterlagen nicht entnommen werden. Jedoch erhielten gemäß Studienbericht alle 4 Teilnehmende innerhalb der angegebenen Dosistitration eine Dosisescalation auf 600 µg/kg BID.

*Tabelle 13: Allgemeine Angaben; Studie RISE (Interim-Datenschnitt: 23.06.2022)*

<b>Studie RISE Allgemeine Angaben</b>	<b>MRX</b>
ITT-Population <sup>1)</sup> , n (%)	4 (100)
Sicherheitspopulation <sup>1)</sup> , n (%)	4 (100)
Abbruch der Einnahme der Studienmedikation, n (%)	0
<i>Studienphase, n (%)</i>	
Kernstudie <sup>2)</sup>	0
Langfristige Behandlungsphase <sup>3)</sup>	4 (100)
Abgeschlossene Behandlung	0
Mittlere Behandlungsdauer, Tage (SD) <sup>4)</sup>	121,0 (34,0)
Mediane Behandlungsdauer, Tage (min; max)	114,5 (93; 162)
Mediane Beobachtungsdauer	k. A.
Durchschnittliche Dosis (mg/Tag)	k. A.

<sup>1)</sup> Definition siehe Kapitel 2.4.

<sup>2)</sup> Bestehend aus der Dosisescalation (Woche 1–6) und der stabilen Phase (Woche 6–13).

<sup>3)</sup> Ab Woche 13.

<sup>4)</sup> Behandlungsdauer = Datum der letzten Dosis des Studienmedikaments - Datum der ersten Dosis des Studienmedikaments + 1 Tag.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; MRX: Maralixibat; PFIC: Progressive familiäre intrahepatische Cholestase; SD: Standardabweichung.

Die Studie MARCH-PFIC schloss Teilnehmende bis zu einem Alter von 18 Jahren ein, wobei die Patientinnen und Patienten mehrheitlich unter 6 Jahre alt waren. In der PFIC-Population lag der Altersmedian in beiden Behandlungsgruppen bei 3 Jahren; für die Gesamtpopulation betrug der Altersmedian in der Interventionsgruppe ebenfalls 3 Jahre und in der Kontrollgruppe 3,6 Jahre. In Bezug auf das Geschlecht ergaben sich für beide Populationen keine erheblichen Unterschiede zwischen der Maralixibat- und der Kontrollgruppe. Während die mediane Krankheitsdauer zum Zeitpunkt der Randomisierung in der Gesamtpopulation zwischen den Behandlungsgruppen ähnlich war (Median = 28 Monate für MRX; Median = 29 Monate für die Kontrollen), zeigten sich in der PFIC-Population mit 12 Monaten in der Kontrollgruppe und 28 Monaten in der Interventionsgruppe deutliche Unterschiede.

In Bezug auf Körpergröße und Körpergewicht liegen die z-Scores grundsätzlich unter den geschlechts- und altersbezogenen Normwerten, was auf Wachstumsdefizite hinweist. Nur der Median und Mittelwert für den BMI in der Interventionsgruppe lag sowohl in der PFIC- als auch in der Gesamtpopulation leicht über der geschlechts- und altersbezogenen Norm.

Bei ca. der Hälfte der Gesamtpopulation der Studie MARCH-PFIC wurde PFIC2 diagnostiziert, davon insgesamt 9 Kinder mit t-PFIC2, 5 (10,6 %) in der Behandlungs- und 4 (8,7 %) in der Kontrollgruppe. Eine heterozygote Variante ist lediglich bei 2 Kindern aus der Kontrollgruppe dokumentiert. Für

3 Personen (6,4 %) aus der Maralixibat-Gruppe und 5 (10,9%) der Placebo-Gruppe wurde keine bekannte Variante gefunden. Es wurde keine Person mit PFIC5 in die Studie eingeschlossen. In der Gesamtpopulation hatten 8 Personen bereits eine Operation erhalten (5 im Maralixibat-Arm, 3 im Placebo-Arm) und jeweils 1 Person aus jedem Behandlungsarm wies fluktuierende sBA-Werte auf. Über 80 % der Studienpopulation gab zu Baseline an Ursodeoxycholsäure (UDCA) einzunehmen und gut die Hälfte nahm Rifampicin als Begleitmedikation ein. In Bezug auf die Baseline-Werte von ItchRO, „Klinische Kratzskala“ und sBA liegen keine erheblichen Unterschiede zwischen den Behandlungsgruppen vor.

In der Studie RISE lag der Altersmedian bei 8,5 Monaten, wobei 2 der insgesamt 4 Kinder männlich und 2 weiblich waren. Auch in dieser Studie zeigten sich z-Scores für Körpergewicht und Körpergröße, die unter den alters- und geschlechtsspezifischen Normwerten lagen. Bei 2 Kindern ist als Genotyp PFIC2 angegeben und jeweils 1 Kind hatte PFIC1 bzw. PFIC4. Informationen zu der regionalen Herkunft der Kinder oder dem betreuerberichteten Juckreiz (ItchRO(Obs)) liegen nicht vor.

Tabelle 14: Charakterisierung der Studienpopulationen der Studien MARCH-PFIC und RISE

Charakterisierung der Studienpopulationen	Studie MARCH-PFIC				Studie RISE MRX N = 4
	PFIC-Population		Gesamtpopulation		
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46	
<i>Alter (Jahre)</i>					
MW (SD)	4,9 (4,10)	4,4 (3,61)	4,8 (4,15)	4,7 (3,57)	7,8 (3,59) <sup>1)</sup>
Median (min; max)	3,0 (1; 15)	3,0 (1; 13)	3,0 (1; 17)	3,5 (1; 14)	8,5 (3; 11) <sup>1)</sup>
<i>Altersgruppe, n (%)</i>					
≤ 1 Jahr	0	0	0	0	4 (100)
≥ 1 bis < 6 Jahre	22 (66,7)	19 (61,3)	32 (68,1)	29 (63,0)	0
≥ 6 bis < 13 Jahre	8 (24,2)	11 (35,5)	11 (23,4)	15 (32,6)	0
≥ 13 bis ≤ 18 Jahre	3 (9,1)	1 (3,2)	4 (8,5)	2 (4,4)	0
<i>Geschlecht, n (%)</i>					
männlich	17 (51,5)	13 (41,9)	20 (42,6)	22 (47,8)	2 (50,0)
weiblich	16 (48,5)	18 (58,1)	27 (57,4)	24 (52,2)	2 (50,0)
<i>Abstammung (genetisch), n (%)</i>					
amerikanisch oder Alaska-Ureinwohner	3 (9,1)	4 (12,9)	3 (6,4)	4 (8,7)	0
asiatisch	3 (9,1)	0	3 (6,4)	0	0
afroamerikanisch	1 (3,0)	2 (6,5)	1 (2,1)	2 (4,4)	0
kaukasisch/weiß	24 (72,7)	19 (61,3)	36 (76,6)	34 (73,9)	3 (75,0)
mehr als eine Herkunft berichtet	2 (6,1)	4 (12,9)	3 (6,4)	4 (8,7)	0
nicht berichtet	0	2 (6,5)	1 (2,1)	2 (4,4)	1 (25,0)
<i>Ethnie, n (%)</i>					
hispanisch oder lateinamerikanisch	16 (48,5)	13 (41,9)	18 (38,3)	17 (37,0)	0
nicht hispanisch oder lateinamerikanisch	17 (51,5)	16 (51,6)	28 (59,6)	27 (58,7)	3 (75,0)
nicht berichtet	0	2 (6,5)	1 (2,1)	2 (4,4)	1 (25,0)
<i>Region, n (%)</i>					
Asien	2 (6,1)	0	2 (4,3)	0	k. A.
Europa	7 (21,2)	7 (22,6)	14 (29,8)	10 (21,7)	k. A.
Mittlerer Osten	5 (15,2)	5 (16,1)	7 (14,9)	9 (19,6)	k. A.
Nordamerika	9 (27,3)	12 (38,7)	13 (27,7)	19 (41,3)	k. A.
Süd- und Mittelamerika	10 (30,3)	7 (22,6)	11 (23,4)	8 (17,4)	k. A.

Charakterisierung der Studienpopulationen	Studie MARCH-PFIC				Studie RISE
	PFIC-Population		Gesamtpopulation		MRX N = 4
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46	
<i>Körpergröße (z-score)</i>					
MW (SD)	-2,08 (1,29)	-2,06 (1,48)	-2,01 (1,33)	-1,91 (1,32)	-1,46 (1,29)
Median (min; max)	-1,96 (-4,92; 0,67)	1,48 (-4,54; 0,88)	-1,96 (-4,92; 0,67)	-1,97 (-4,54; 0,88)	-1,29 (-2,9; -0,4)
<i>Körpergewicht (z-score)</i>					
MW (SD)	-1,75 (1,29)	-1,28 (1,33)	-1,56 (1,36)	-1,22 (1,22)	-1,11 (0,60)
Median (min; max)	-1,46 (-5,08; 1,01)	-1,09 (-3,88; 1,46)	-1,4 (-5,08; 1,08)	-1,14 (-3,88; 1,46)	-1,16 (-1,8; -0,4)
<i>BMI (z-score)</i>					k. A.
MW (SD)	-0,45 (1,09)	0,14 (1,13)	-0,26 (1,11)	0,07 (1,16)	k. A.
Median (min; max)	-0,53 (-3,55; 1,87)	0,11 (-2,64; 2,34)	-0,47 (-3,55; 2,14)	0,12 (-2,64; 2,34)	k. A.
<i>Ausgewählte Begleittherapie zu Baseline, n (%)<sup>2)</sup></i>					
UDCA	27 (81,8)	30 (96,8)	39 (83,0)	39 (84,8)	k. A.
Rifampicin	15 (54,5)	15 (48,4)	26 (55,3)	23 (50,0)	k. A.
<i>Zeit ab Diagnose bis Randomisierung (Monate)</i>					
Median (min; max)	28,0 (2; 136)	12,0 (1; 109)	28,0 (1; 136)	29,0 (1; 110)	k. A.
<i>PFIC-Typ (Genotyp), n (%)<sup>3)</sup></i>					
PFIC2 (ABCB11)	14 (42,4)	17 (54,8)	23 (48,9)	22 (47,8)	2 (50,0)
nt-PFIC2 <sup>4)</sup>	14 (42,4)	17 (54,8)	18 (38,3)	18 (39,1)	0
BSEP1	2 (6,1)	6 (19,4)	4 (8,5)	6 (13,0)	0
BSEP2	12 (36,4)	11 (35,5)	14 (29,8)	12 (26,1)	0
t-PFIC2 (BSEP3)	0	0	5 (10,6)	4 (8,7)	0
PFIC1 (ATP8B1) <sup>5)</sup>	7 (21,2)	6 (19,4)	9 (19,1)	8 (17,4)	1 (25,0)
PFIC3 (ABCB4)	4 (12,1)	5 (16,1)	4 (8,5)	5 (10,9)	0
PFIC4 (TJP2) <sup>6)</sup>	6 (18,2)	1 (3,2)	6 (12,8)	2 (4,4)	1 (25,0)
PFIC6 (MYO5B)	2 (6,1)	2 (6,5)	2 (4,3)	2 (4,4)	0
Heterozygote Variante	0	0	0	2 (4,4)	0
Keine Variante gefunden	0	0	3 (6,4)	5 (10,9)	0
<i>Detektierte Lebermasse zu Baseline, n (%)</i>					
Ja	1 (3,0)	0	1 (2,1)	0	k. A.
Nein	32 (97,0)	31 (100)	46 (97,9)	46 (100)	k. A.

Charakterisierung der Studienpopulationen	Studie MARCH-PFIC				Studie RISE MRX N = 4
	PFIC-Population		Gesamtpopulation		
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46	
<i>Score „Klinische Kratzskala“</i>					
n (%)	33 (100)	31 (100)	47 (100)	46 (100)	4 (100)
MW (SD)	2,8 (0,74)	2,6 (0,99)	2,7 (0,85)	2,7 (0,95)	2,0 (0,82)
Median (min; max)	3,0 (1; 4)	3,0 (1; 4)	3,0 (0; 4)	3,0 (1; 4)	2,0 (1; 3)
<i>Morgendlicher ItchRO(Obs), 4-wöchentlicher mittlerer Schweregrad-Score<sup>7)</sup></i>					
n (%)	33 (100)	31 (100)	47 (100)	46 (100)	k. A.
MW (SD)	2,85 (0,88)	2,73 (0,88)	2,85 (0,83)	2,93 (0,84)	k. A.
Median (min; max)	2,93 (1,00; 4,00)	2,57 (1,67; 4,00)	2,93 (1,00; 4,00)	2,96 (1,67; 4,00)	k. A.
<i>Morgendlicher ItchRO(Pt), 4-wöchentlicher mittlerer Schweregrad-Score</i>					
n (%)	5 (15,2)	4 (12,9)	7 (14,9)	6 (13,0)	k. A.
MW (SD)	3,16 (1,06)	1,52 (0,25)	3,04 (0,99)	2,09 (1,03)	k. A.
Median (min; max)	3,49 (1,32; 4,00)	1,63 (1,14; 1,67)	3,49 (1,32; 4,00)	1,66 (1,14; 4,00)	k. A.
<i>Gesamt-sBA (µmol/l)<sup>8)</sup></i>					
n (%)	31 (93,9)	31 (100)	45 (95,7)	46 (100)	4 (100)
MW (SD)	254,33 (139,52)	272,3 (147,40)	262,62 (150,72)	243,31 (152,94)	211,72 (117,13)
Median (min; max)	246,03 (4,25; 503,53)	262,09 (2,10; 549,20)	254,88 (4,25; 536,39)	226,46 (2,10; 549,20)	220,34 (60,9; 345,3)
<i>Gesamtbilirubin (mg/d)</i>					
n (%)	33 (100)	31 (100)	47 (100)	46 (100)	4 (100)
MW (SD)	4,12 (3,80)	4,04 (4,46)	4,1 (3,74)	3,8 (4,64)	45,94 (31,60) <sup>9)</sup>
Median (min; max)	2,8 (0,30; 14,65)	2,55 (0,25; 17,00)	2,8 (0,30; 14,65)	2,38 (0,25; 21,40)	45,79 (12,9; 79,3) <sup>6)</sup>
<i>Direktes Bilirubin (mg/d)</i>					
n (%)	33 (100)	31 (100)	47 (100)	46 (100)	4 (100)
MW (SD)	2,98 (2,82)	2,93 (3,35)	2,99 (2,80)	2,77 (3,52)	34,86 (26,17) <sup>6)</sup>
Median (min; max)	2,05 (0,15; 10,35)	1,85 (0,10; 12,65)	2,05 (0,15; 10,35)	1,70 (0,10; 16,40)	32,97 (8,70; 64,8) <sup>6)</sup>

- 1) Alter in Monaten.
- 2) Behandlungen, die vor der ersten Dosis des Studienmedikaments begonnen wurden, gelten als vorherige Behandlungen, unabhängig davon, ob sie vor der ersten Dosis des Studienmedikaments beendet wurden oder nicht. Jede Behandlung, die nach der ersten Dosis des Studienmedikaments fortgesetzt oder begonnen wird, wird als Begleitbehandlung betrachtet. Wenn eine Behandlung vor der ersten Dosis des Studienmedikaments beginnt und nach der ersten Dosis des Studienmedikaments fortgesetzt wird, gilt das Medikament sowohl als vorherige als auch als begleitende Behandlung.
- 3) Mutationsstatus u. a. berücksichtigt als Stratifizierungsfaktor (siehe Tabelle 2).
- 4) Jeweils 1 Person aus den Behandlungsgruppen der Gesamtpopulation hatte fluktuierende sBA-Werte und 3 weitere Personen aus der Maralixibat-Gruppe wurden bereits einer Operation unterzogen.
- 5) Jeweils 2 Person aus den Behandlungsgruppen der Gesamtpopulation wurden bereits einer Operation unterzogen.
- 6) Eine Person der Placebo-Gruppe wurde bereits einer Operation unterzogen.
- 7) Zentrales Kriterium für Einschluss in die Studie: Mittlerer morgendlicher ItchRO(Obs)-Score  $\geq 1,5$  für 4 aufeinanderfolgende Wochen in der Screeningphase bis zu Baseline.
- 8) Diagnostisches Kriterium einer Cholestase und Einschlusskriterium: Gesamt-sBA  $\geq 3 \times$  ULN.
- 9) Einheit:  $\mu\text{mol/l}$ .

Abkürzungen: BSEP: Gallensalz-Exportpumpe; ItchRO(Obs): Itch Reported Outcome (Observer Version); ItchRO(Pt): Itch Reported Outcome (Patient Version); k. A.: keine Angabe; MRX: Maralixibat; MW: Mittelwert; nt-PFIC: non-truncated PFIC; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; sBA: Serum-Gallensäure; SD: Standardabweichung; t-PFIC: truncated PFIC; UDCA: Ursodeoxycholsäure; ULN: Upper Limit of Normal.

### **Protokollverletzungen**

Anhand der in den Studien MARCH-PFIC und RISE dokumentierten Protokollverletzungen ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

### **Begleitmedikation**

In der PFIC-Population der Studie MARCH-PFIC wurden Antihämorrhagika öfter unter Placebo (71 %) als in der Maralixibat-Gruppe (52 %) eingenommen. Auch UDCA war verbreiteter im Placebo-Arm (82 %) verglichen mit dem Maralixibat-Arm (97 %). Dagegen wurden systemische Antihistaminika häufiger unter Maralixibat (58 %) als unter Placebo (42 %) eingenommen. Anhand der in der PFIC-Population dokumentierten Begleitmedikation ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens. Für die Gesamtpopulation zeigen sich keine relevanten Unterschiede zwischen den Behandlungsarmen.

Angaben zu Begleitmedikationen für die Studie RISE liegen nicht vor.

### **Folgetherapien**

In der Studie MARCH-PFIC sind keine Folgetherapien dokumentiert, ein Großteil der Patientinnen und Patienten ging in die Extensionsstudie MRX-503 über. Die Studie RISE ist laufend und alle Patientinnen und Patienten befanden sich zum Datenschnitt noch unter Behandlung.

## **3.2 Mortalität**

In der Studie MARCH-PFIC sind zum finalen Datenschnitt keine Todesfälle aufgetreten, genau wie in der Studie RISE zum Interim-Datenschnitt vom 23.06.2022.

## **3.3 Morbidität**

### **Pruritus mittels ItchRO**

Angaben zur Anzahl fehlender Werte bzw. dem Anteil der fehlenden Tagebucheinträge zur Berechnung der Rücklaufquote konnten den Studienunterlagen weder für den ItchRO(Obs) noch für den ItchRO(Pt) entnommen werden. Im Rahmen einer Post-hoc-Analyse / Im Dossier gibt der pU an, dass die Patientinnen und Patienten der Interventionsgruppe kumuliert insgesamt 6.126 Tage im gesamten Beobachtungszeitraum waren, davon ist an 5.562 Tagen (90,8 %) eine Erhebung des ItchRO dokumentiert. Für die Kontrollgruppe werden 5.311 Tage im Beobachtungszeitraum berichtet, mit insgesamt 4.970 Tagen (93,6 %) mit ItchRO-Erhebungen. Aus diesen Angaben lässt sich eine ausreichende Datenvollständigkeit für den ItchRO(Obs) ableiten. Es ist jedoch zu beachten, dass keine konkreten Rücklaufquoten für die einzelnen Analysen des ItchRO abgeleitet werden können.

Die ItchRO-Responderanalysen beziehen sich ausschließlich auf die Erhebung am Morgen durch die betreuende Person. In der Nutzenbewertung wird insbesondere der Anteil der Patientinnen und Patienten mit einer Verbesserung  $\geq 1$  Punkt oder mit einem Schweregrad-Score  $\leq 1$  zu Woche 15–26 als relevant erachtet. Dabei zeigten sich signifikant mehr Responder in der Interventionsgruppe im Vergleich zur Kontrollgruppe. In einer weiteren Auswertung der Anzahl der Tage mit einem morgendlichen ItchRO(Obs)-Schweregrad  $\leq 1$  für die letzten 84 Tage der Studie (entsprechend Woche 15–26) hatten Teilnehmende im Maralixibat-Arm im Mittel an 62,5 Tagen (SD: 24,7) einen Schweregrad  $\leq 1$ , Teilnehmende im Placebo-Arm durchschnittlich an 39,5 Tagen (SD: 29,7). Dieser Unterschied war statistisch signifikant (MWD: 23,0 (95%-KI: [6,53; 39,4]));  $p = 0,0075$ ).

**Tabelle 15: Responderanalyse zur Verbesserung im ItchRO(Obs) in der Erhebung am Morgen;  
Studie MARCH-PFIC, ITT-Population**

Studie MARCH-PFIC ItchRO(Obs) <sup>1)</sup>	PFIC-Population			Gesamtpopulation	
	MRX N = 33	PBO N = 31	RR [95%-KI]; p-Wert <sup>2)</sup>	MRX N = 47	PBO N = 46
Anzahl fehlender Werte <sup>3)</sup>	k. A.	k. A.	-	k. A.	k. A.
Verbesserung $\geq 1$ Punkt o. Schweregrad-Score $\leq 1$ zu Woche 15–26, n (%) <sup>4)</sup>	21 (63,6)	8 (25,8)	2,47 [1,29; 4,72]; 0,0023	k. A.	k. A.

<sup>1)</sup> Skala von 0 bis 4; höhere Werte entsprechen einer höheren Belastung.

<sup>2)</sup> P-Wert berechnet mittels Barnard's exaktem Test.

<sup>3)</sup> Keine Angabe zur Anzahl fehlender Werte in der Analyse. Kumuliert für alle Personen ist eine Erhebung an 5.562 Tagen (90,8 %) für den ItchRO(Obs) in der Interventionsgruppe und an 4.970 Tagen (93,6 %) in der Kontrollgruppe dokumentiert, woraus sich eine hohe Datenvollständigkeit für den ItchRO(Obs) ableiten lässt.

<sup>4)</sup> Verwendet wurde der mittlere Schweregrad der drei 4-Wochen-Perioden (Wochen 15–18, 19–22 und 23–26). Eine Person wurde als Non-Responder klassifiziert, wenn der mittlere 4-wöchentliche Score zu Baseline oder alle drei 4-Wochen-Perioden fehlten.

Abkürzungen: ItchRO(Obs): Itch Reported Outcome (Observer Version); ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MRX: Maralixibat; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; RR: Relatives Risiko.

Zusätzlich zu den Responderanalysen werden ergänzend kontinuierliche Analysen der mittleren Veränderung dargestellt. Für alle 3 Auswertungen zeigte sich für den Zeitraum von Woche 15 bis Woche 26 eine signifikante Verbesserung des Pruritus-Schweregrads gegenüber Baseline. Die Effekte sind anhand des Hedges' g als klinisch relevant einzustufen.

Die vorliegenden Sensitivitätsanalysen in Bezug auf fehlende Werte (MAR-Ansatz und multiple Imputation) für die Veränderung des morgendlichen ItchRO-Werts, weisen auf die Robustheit der Ergebnisse hin.

In der Studie RISE wurde der ItchRO(Obs) als explorativer Endpunkt definiert, im Interim-Studienbericht werden jedoch keine Ergebnisse zu diesem Endpunkt berichtet.

**Tabelle 16: Mittlere Veränderung im Schweregrad des ItchRO(Obs) zu Woche 15–26 gegenüber Baseline; Studie MARCH-PFIC, ITT-Population (ergänzend dargestellt)**

Studie MARCH-PFIC ItchRO(Obs) <sup>1)</sup>	PFIC-Population		Gesamtpopulation	
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46
<b>Morgendlicher ItchRO(Obs)</b>				
<i>Baseline</i>				
Rücklauf, n (%)	33 (100)	31 (100)	k. A.	k. A.
MW (SD)	2,85 (0,88)	2,73 (0,88)		
<i>Veränderung zu Woche 15–26<sup>2)</sup></i>				
Rücklauf, n (%)	k. A. <sup>3)</sup>	k. A. <sup>3)</sup>	k. A.	k. A.
LS-MW (SE)	-1,81 (0,18)	-0,61 (0,19)		
LS-MWD [95%-KI] <sup>2)</sup>	-1,20 [-1,73; -0,67]			
p-Wert	< 0,0001			k. A.
SMD [95%-KI] <sup>4)</sup>	-1,13 [-1,66; -0,60]			

Studie MARCH-PFIC ItchRO(Obs) <sup>1)</sup>	PFIC-Population		Gesamtpopulation	
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46
<b>Abendlicher ItchRO(Obs)</b>				
<i>Baseline</i>				
Rücklauf, n (%)	33 (100)	31 (100)	k. A.	k. A.
MW (SD)	2,79 (0,82)	2,66 (0,89)		
<i>Veränderung zu Woche 15–26<sup>2)</sup></i>				
Rücklauf, n (%)	k. A. <sup>3)</sup>	k. A. <sup>3)</sup>	k. A.	k. A.
LS-MW (SE)	-1,77 (0,18)	-0,61 (0,20)		
LS-MWD [95%-KI] <sup>2)</sup>	-1,16 [-1,69; -0,63]			
p-Wert	< 0,0001			
SMD [95%-KI] <sup>4)</sup>	-1,08 [-1,61; -0,56]			
<b>Maximum aus Abend und Morgen ItchRO(Obs)</b>				
<i>Baseline</i>				
Rücklauf, n (%)	33 (100)	31 (100)	k. A.	k. A.
MW (SD)	3,01 (0,82)	2,88 (0,83)		
<i>Veränderung zu Woche 15–26<sup>2)</sup></i>				
Rücklauf, n (%)	k. A. <sup>3)</sup>	k. A. <sup>3)</sup>	k. A.	k. A.
LS-MW (SE)	-1,89 (0,19)	-0,69 (0,20)		
LS-MWD [95%-KI] <sup>2)</sup>	-1,20 [-1,74; -0,66]			
p-Wert	< 0,0001			
SMD [95%-KI] <sup>4)</sup>	-1,10 [-1,63; -0,58]			

<sup>1)</sup> Skala von 0 bis 4; höhere Werte entsprechen einer höheren Belastung.

<sup>2)</sup> Schätzer aus einem MMRM mit „Veränderung zu Baseline“ als abhängiger Variable und den festen kategoriellen Effekten „Behandlungsgruppe“, „Zeitpunkt“, „PFIC-Typ“ und „Interaktion zwischen Behandlungsgruppe und Zeitpunkt“ und den kontinuierlichen festen Kovariaten „Mittlerer Baseline-Score“ und „Interaktion zwischen Baseline-Score und Zeitpunkt“.

<sup>3)</sup> Keine Angabe zur Anzahl fehlender Werte in der Analyse. Kumuliert für alle Personen ist eine Erhebung an 5.562 Tagen (90,8 %) für den ItchRO(Obs) in der Interventionsgruppe und an 4.970 Tagen (93,6 %) in der Kontrollgruppe dokumentiert, woraus sich eine hohe Datenvollständigkeit für den ItchRO(Obs) ableiten lässt.

<sup>4)</sup> SMD nach Hedges' g.

Abkürzungen: ItchRO(Obs): Itch Reported Outcome (Observer Version); ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Kleinste Quadrate; MMRM: Mixed Model for Repeated Measures; MRX: Maralixibat; MW: Mittelwert; MWD: Mittelwertdifferenz; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; SD: Standardabweichung, SE: Standardfehler; SMD: Standardisierte Mittelwertdifferenz.

Der ItchRO(Pt) sollte in der Studie MARCH-PFIC von allen Personen ab einschließlich 9 Jahren erhoben werden. Aus den Studienunterlagen geht nicht hervor, wie groß diese Teilpopulation ist. Zu Baseline liegen lediglich Daten für 5 Teilnehmende aus der Maralixibat-Gruppe und für 4 aus der Placebo-Gruppe vor. Für den ItchRO(Pt) waren nur kontinuierliche Analysen bezüglich der Veränderung des Schweregrads geplant, Responderanalysen waren nicht vorgesehen. Es werden zudem keine Angaben bezüglich fehlender Werte und Rücklaufquoten gemacht, weshalb die Ergebnisse des ItchRO(Pt) nicht dargestellt werden.

Für die Studie RISE liegen keine Ergebnisse zum ItchRO vor.

### Klinische Kratzskala (ergänzend dargestellt)

Für die „Klinische Kratzskala“ liegen in der Studie MARCH-PFIC ausschließlich Ergebnisse für die mittlere Veränderung des Zeitraums von Woche 18 bis Woche 26 gegenüber Baseline vor. Angaben zu fehlenden Werten oder der Rücklaufquote konnten den Unterlagen des pU nicht entnommen werden. Die Ergebnisse des Endpunkts können daher nicht berücksichtigt werden.

In der Studie RISE erfolgte die Auswertung der klinischen Kratzskala deskriptiv, analog zur Studie MARCH-PFIC wurden keine Responderanalysen definiert. Die Ergebnisse beziehen sich auf die Kernstudie, in welcher die „Klinische Kratzskala“ zu Baseline, Woche 6 und Woche 13 erhoben wurde. Während sich zu Woche 6 im Vergleich zu Baseline eine leichte Reduktion um -0,3 Punkte ergab, stieg der Wert bis Woche 13 wieder an und lag damit am Ende der Kernstudie 0,3 Punkte über dem Baseline-Wert. Die mediane Veränderung zu Woche 6 liegt bei 0, steigt dann aber bis Woche 13 auf 0,5 an. Die Standardabweichungen sowie die Werte für Minimum und Maximum zu den verschiedenen Erhebungszeitpunkten deuten darauf hin, dass sich die Wirksamkeit von Maralixibat in Bezug auf die Reduktion des Pruritus im Laufe der Zeit zunehmend zwischen den Patientinnen und Patienten unterscheidet.

*Tabelle 17: Veränderung des Scores der „Klinischen Kratzskala“ im Vergleich zu Baseline; Studie RISE, Sicherheitspopulation*

Studie RISE Klinische Kratzskala <sup>1)</sup>	MRX N = 4	
	Beobachteter Wert	Veränderung gegenüber Baseline
<i>Baseline</i> <sup>2)</sup>		
Rücklauf, n (%)	4 (100)	-
MW (SD)	2,0 (0,82)	
Median (min; max)	2,0 (1; 3)	
<i>Woche 6</i>		
Rücklauf, n (%)	4 (100)	4 (100)
MW (SD)	1,8 (1,26)	-0,3 (0,50)
Median (min; max)	2,0 (0; 3)	0,0 (-1; 0)
<i>Woche 13</i>		
Rücklauf, n (%)	4 (100)	4 (100)
MW (SD)	2,3 (1,71)	0,3 (0,96)
Median (min; max)	2,5 (0; 4)	0,5 (-1; 1)

<sup>1)</sup> Werte von 0 bis 4; höhere Werte entsprechen einer höheren Belastung.

<sup>2)</sup> Mittelwert der letzten 2 nicht-fehlenden Messungen vor der ersten Dosis von Maralixibat oder letzte Messung vor der ersten Dosis, wenn nur ein Wert vorlag.

Abkürzungen: MRX: Maralixibat; MW: Mittelwert; SD: Standardabweichung.

### Schlafstörungen mittels EDQ(Obs)

In der Studie MARCH-PFIC füllten Patientinnen und Patienten ab dem Alter von 9 Jahren den EDQ(Pt) aus, wofür jedoch keine Ergebnisse vorliegen. Die Erhebung und Auswertungen des EDQ(Obs) beziehen sich lediglich auf die Teilpopulation der unter 9-Jährigen. Es ist unklar, wie groß die Teilpopulation der unter 9-Jährigen in der jeweiligen Studienpopulation ist. Zu Baseline lagen Werte für 27 Personen der Maralixibat-Gruppe und 26 der Placebo-Gruppe vor. Angaben zur Anzahl der Tagebucheinträge bzw. zu den Rücklaufquoten konnten für die Analysen des EDQ(Obs) nicht identifiziert werden. Aus diesem Grund werden die Ergebnisse für den EDQ(Obs) nicht dargestellt.

In der Studie RISE wurde der EDQ(Obs) nicht erhoben.

## Körperliche Entwicklung

Für die Analyse zu Woche 18–26 im Vergleich zu Baseline lagen sowohl für die Körpergröße als auch das Körpergewicht Daten von 32 Personen (97 %) im Interventionsarm und von 28 (90 %) im Kontrollarm vor. Auch für die Auswertungen zum Wachstum ist unklar, wie viele Personen zum jeweiligen Zeitpunkt in die Analyse eingehen. Im Vergleich zu Baseline zeigte sich für die Körpergröße kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen. Hinsichtlich des Körpergewichts (z-Score) zeigte sich eine leichte Zunahme für beide Behandlungsgruppen, wobei diese in der Maralixibat-Gruppe größer war. Der Effekt war statistisch signifikant jedoch nicht klinisch relevant gemäß Hedges' g. Es ist zu berücksichtigen, dass die Kinder der Maralixibat-Gruppe zu Baseline mit -1,75 stärker von der Norm abwichen im Vergleich zu Placebo -1,28.

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

*Tabelle 18: Körperliche Entwicklung anhand Körpergröße (z-Score) und Körpergewicht (z-Score) zu Woche 18–26 gegenüber Baseline; Studie MARCH-PFIC, ITT-Population*

Studie MARCH-PFIC Körperliche Entwicklung	PFIC-Population		Gesamtpopulation	
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46
<b>Körpergröße (z-Score)</b>				
<i>Baseline</i> n (%) MW (SD)	32 (97,0) -2,08 (1,31)	31 (100) -2,06 (1,48)	k. A.	k. A.
<i>Veränderung zu Woche 18–26<sup>1)</sup></i> n (%) LS-MW (SE)	32 (97,0) 0,08 (0,09)	28 (90) -0,13 (0,09)	k. A.	k. A.
LS-MWD [95%-KI] <sup>1)</sup> p-Wert	0,21 [-0,04; 0,45] 0,09		k. A.	
<b>Körpergewicht (z-Score)</b>				
<i>Baseline</i> n (%) MW (SD)	32 (97,0) -1,75 (1,31)	31 (100) -1,28 (1,33)	k. A.	k. A.
<i>Veränderung zu Woche 18–26<sup>1)</sup></i> n (%) LS-MW (SE)	32 (97,0) 0,35 (0,07)	28 (90) 0,12 (0,08)	k. A.	k. A.
LS-MWD [95%-KI] <sup>1)</sup> p-Wert SMD [95%-KI] <sup>2)</sup>	0,23 [0,01; 0,44] 0,039 0,53 [0,02; 1,03]		k. A.	

<sup>1)</sup> Durchschnitt der Wochen 18, 22 und 26 aus dem MMRM als gleich gewichteter Durchschnitt der 3 individuellen Visiten-spezifischen Schätzer. Schätzer aus einem MMRM mit „Veränderung zu Baseline“ als abhängiger Variable und den festen kategoriellen Effekten „Behandlungsgruppe“, „Zeitpunkt“, „PFIC-Typ“ und „Interaktion zwischen Behandlungsgruppe und Zeitpunkt“ und den kontinuierlichen festen Kovariaten „Mittlerer Baseline-Score“ und „Interaktion zwischen Baseline-Score und Zeitpunkt“.

<sup>2)</sup> SMD nach Hedges' g.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Kleinste Quadrate; MMRM: Mixed Model for Repeated Measures; MRX: Maralixibat; MW: Mittelwert; MWD: Mittelwertdifferenz; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; SD: Standardabweichung, SE: Standardfehler; SMD: Standardisierte Mittelwertdifferenz.

### **PedsQL-Fatigue**

In der Studie MARCH-PFIC wurde die elternberichtete PedsQL-Fatigue für Kinder ab dem Alter von 2 Jahren ausgefüllt. Es ist unklar, wie groß die Teilpopulation der  $\geq 2$ -Jährigen ist. Ergebnisse für die von den Kindern selbst eingeschätzten Versionen des Fragebogens konnten nicht identifiziert werden, genau wie Angaben zu den Rücklaufquoten bzw. der Anzahl der beantworteten Fragebögen zum jeweiligen Erhebungszeitpunkt. Aus diesem Grund erfolgt keine Darstellung der Ergebnisse für die PedsQL-Fatigue.

In der Studie RISE wurde dieser Endpunkt nicht erhoben.

### **Leberassoziierte Ereignisse (ergänzend dargestellt)**

In der Studie MARCH-PFIC sind zum finalen Datenschnitt in der PFIC-Population keine leberassoziierten Ereignisse dokumentiert.

Der Endpunkt wurde in der Studie RISE nicht erhoben.

## **3.4 Lebensqualität**

Für die Studie MARCH-PFIC liegen ausschließlich Ergebnisse zum Gesamtscore für die elternberichteten Versionen des PedsQL in der PFIC-Population vor. Ergebnisse für die von den Kindern selbst eingeschätzten Versionen des Fragebogens konnten nicht identifiziert werden, genau wie Angaben zu den Rücklaufquoten bzw. der Anzahl der beantworteten Fragebögen zum jeweiligen Erhebungszeitpunkt. Die Ergebnisse für den Gesamtscore des PedsQL werden daher nicht dargestellt.

## **3.5 Sicherheit**

Angaben zur Beobachtungsdauer in der Studie MARCH-PFIC liegen nicht vor. Hinsichtlich der Behandlungsdauer liegen Ergebnisse nur für den Maralixibat-Arm vor, die im Median sowohl für die PFIC-Population als auch die Gesamtpopulation 183 Tage beträgt. Aus Angaben zur Datenvollständigkeit des ItchRO(Obs) geht eine vergleichbare Beobachtungsdauer in den beiden Studienarmen hervor.

Bezogen auf die Gesamtpopulation trat bei allen Teilnehmenden in der Maralixibat-Gruppe sowie bei ca. 94 % der Teilnehmenden in der Placebo-Gruppe mindestens ein UE auf. Grundsätzlich sind für keine der beiden Populationen erhebliche Unterschiede zwischen den Behandlungsgruppen zu erkennen und der Anteil an schweren UE oder SUE war generell gering. Lediglich 1 Person im Maralixibat-Arm brach die Studie aufgrund eines UE ab.

**Tabelle 19: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase; Studie MARCH-PFIC, Sicherheitspopulation**

Studie MARCH-PFIC Zusammenfassung der UE <i>Personen mit mind. einem ...</i>	PFIC-Population		RR [95%-KI]; p-Wert	Gesamtpopulation		RR [95%-KI]; p-Wert
	MRX N = 33 n (%)	PBO N = 31 n (%)		MRX N = 47 n (%)	PBO N = 46 n (%)	
UE (ergänzend dargestellt)	33 (100)	30 (96,8)	1,03 [0,95; 1,13]; 0,302	47 (100)	43 (93,5)	k. A.
Schwere UE <sup>1)</sup>	1 (3,0)	1 (3,2)	0,94 [0,06; 14,38]; 0,964	3 (6,4)	3 (6,5)	k. A.
SUE	3 (9,1)	3 (9,7)	0,94 [0,20; 4,31]; 0,936	5 (10,6)	3 (6,5)	k. A.
UE, das zum Abbruch der Studienmedikation führte	0	0	0,94 [0,02; 46,04]; n. b.	1 (2,1)	0	k. A.

<sup>1)</sup> In den Studienunterlagen war eine Schweregradeinteilung anhand der Kategorien „mild“, „moderat“ und „schwer“ vorgesehen. Laut Dossier erfolgte jedoch die Einteilung in Schweregrade gemäß CTCAE-Version 4.0. Nähere Angaben zu einer post hoc erfolgten Schweregradeinteilung nach CTCAE liegen nicht vor.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; k. A.: keine Angabe; KI: Konfidenzintervall; MRX: Maralixibat; n. b.: nicht bewertbar; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

In der Studie RISE hatten alle Patientinnen und Patienten mindestens ein UE, nur ein Kind zeigte ein SUE. Kein Kind brach die Studie aufgrund von UE ab.

**Tabelle 20: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der Behandlungsphase; Studie RISE, Sicherheitspopulation**

Studie RISE Zusammenfassung der UE <i>Personen mit mind. einem ...</i>	MRX N = 4 n (%)
UE (ergänzend dargestellt)	4 (100)
UE CTCAE-Grad $\geq 3$ <sup>1)</sup>	0
SUE	1 (25,0)
UE, das zum Abbruch der Studienmedikation führte	0

<sup>1)</sup> Die Beurteilung des Schweregrads in der Studie RISE erfolgte laut Studienprotokoll und SAP anhand CTCAE-Version 5.0, laut Dossier anhand Version 4.0.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; MRX: Maralixibat; SAP: Statistischer Analyseplan; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

### Unerwünschte Ereignisse nach Systemorganklasse und Preferred Term

Statistisch signifikante Unterschiede zwischen den Studienarmen zuungunsten von Maralixibat zeigten sich nur für die SOC „Erkrankungen des Gastrointestinaltrakts“ und den PT „Diarrhö“. Dagegen traten UE der SOC „Leber- und Gallenerkrankungen“ statistisch signifikant häufiger in der Placebo-Gruppe auf. In Bezug auf die SOC „Leber- und Gallenerkrankungen“ ist unklar, ob davon auch Symptome der Grunderkrankung umfasst sein könnten. Auch die PT „Hyperbilirubinämie“ und „Bilirubin im Blut erhöht“ könnten möglicherweise der Grunderkrankung zugeordnet werden, für diese zeigten sich jedoch keine bedeutsamen Gruppenunterschiede.

Tabelle 21: UE mit Inzidenz  $\geq 10\%$ ; Studie MARCH-PFIC, Sicherheitspopulation<sup>1)</sup>

Studie MARCH-PFIC UE	PFIC-Population		RR [95%-KI]; p-Wert	Gesamtpopulation		RR [95%-KI]; p-Wert
	MRX N = 33 n (%)	PBO N = 31 n (%)		MRX N = 47 n (%)	PBO N = 46 n (%)	
<b>MedDRA-Systemorganklasse Preferred Term</b>						
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>16 (48,5)</b>	<b>9 (29,0)</b>	<b>1,67 [0,87; 3,21]; 0,114</b>	<b>18 (38,3)</b>	<b>13 (28,3)</b>	<b>k. A.</b>
Fieber	15 (45,5)	9 (29,0)	1,57 [0,80; 3,05]; 0,178	17 (36,2)	13 (28,3)	k. A.
<b>Erkrankungen der Atem- wege, des Brustraums und Mediastinums</b>	<b>11 (33,3)</b>	<b>6 (19,4)</b>	<b>1,72 [0,72; 4,09]; 0,209</b>	<b>15 (31,9)</b>	<b>10 (21,7)</b>	<b>k. A.</b>
Rhinorrhoe	7 (21,2)	4 (12,9)	1,64 [0,53; 5,07]; 0,3823	8 (17,0)	5 (10,9)	k. A.
Husten	5 (15,2)	3 (9,7)	1,57 [0,41; 6,01]; 0,512	7 (14,9)	5 (10,9)	k. A.
<b>Erkrankungen der Haut und des Unterhautgewebes</b>	<b>8 (24,2)</b>	<b>9 (29,0)</b>	<b>0,84 [0,37; 1,89]; 0,667</b>	<b>10 (21,3)</b>	<b>12 (26,1)</b>	<b>k. A.</b>
Pruritus	4 (12,1)	5 (16,1)	0,75 [0,22; 2,55]; 0,647	5 (10,6)	8 (17,4)	k. A.
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>25 (75,8)</b>	<b>12 (38,7)</b>	<b>1,96 [1,21; 3,17]; 0,003</b>	<b>35 (74,5)</b>	<b>17 (37,0)</b>	<b>k. A.</b>
Abdominalschmerz	8 (24,2)	2 (6,5)	3,76 [0,86; 16,34]; 0,052	10 (21,3)	3 (6,52)	k. A.
Diarrhö	19 (57,6)	5 (16,1)	3,57 [1,52; 8,39]; 0,001	27 (57,4)	9 (19,6)	k. A.
Erbrechen	3 (9,09)	5 (16,1)	0,56 [0,15; 2,16]; 0,399	3 (6,38)	5 (10,9)	k. A.
Obstipation	4 (12,1)	2 (6,5)	1,88 [0,37; 9,54]; 0,440	4 (8,5)	2 (4,3)	k. A.
<b>Infektionen und parasitäre Erkrankungen</b>	<b>18 (54,5)</b>	<b>17 (54,8)</b>	<b>0,99 [0,64; 1,55]; 0,981</b>	<b>26 (55,3)</b>	<b>25 (54,3)</b>	<b>k. A.</b>
Grippe	6 (18,2)	1 (3,2)	5,64 [0,72; 44,20]; 0,057	6 (12,8)	2 (4,3)	k. A.
Nasopharyngitis	4 (12,1)	1 (3,2)	3,76 [0,44; 31,80]; 0,189	5 (10,6)	2 (4,3)	k. A.
Infektion obere Atemwege	-	-	-	3 (6,4)	6 (13,0)	k. A.

Studie MARCH-PFIC UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	PFIC-Population		RR [95%-KI]; p-Wert	Gesamtpopulation		RR [95%-KI]; p-Wert
	MRX N = 33 n (%)	PBO N = 31 n (%)		MRX N = 47 n (%)	PBO N = 46 n (%)	
<b>Leber- und Gallenerkrankungen</b>	<b>1 (3,0)</b>	<b>7 (22,6)</b>	<b>0,13 [0,02; 1,03]; 0,019</b>	<b>2 (4,3)</b>	<b>10 (21,7)</b>	<b>k. A.</b>
Hyperbilirubinämie	-	-	-	0	5 (10,9)	k. A.
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>7 (21,2)</b>	<b>8 (25,8)</b>	<b>0,82 [0,34; 2,00]; 0,667</b>	<b>11 (23,4)</b>	<b>11 (23,9)</b>	<b>k. A.</b>
<b>Untersuchungen</b>	<b>18 (54,5)</b>	<b>14 (45,2)</b>	<b>1,21 [0,73; 1,99]; 0,457</b>	<b>24 (51,1)</b>	<b>18 (39,1)</b>	<b>k. A.</b>
Alanin-Aminotransferase erhöht	4 (12,1)	3 (9,7)	1,25 [0,30; 5,15]; 0,756	6 (12,8)	3 (6,5)	k. A.
Bilirubin im Blut erhöht	5 (15,2)	4 (12,9)	1,17 [0,35; 3,98]; 0,798	7 (14,9)	4 (8,7)	k. A.
<b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>	<b>5 (15,2)</b>	<b>7 (22,6)</b>	<b>0,67 [0,24; 1,89]; 0,450</b>	<b>6 (12,8)</b>	<b>8 (17,4)</b>	<b>k. A.</b>

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

Abkürzungen: -: UE mit Inzidenz < 10 %; k. A.: keine Angabe; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; MRX: Maralixibat; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; PT: Preferred Term; RR: Relatives Risiko; SOC: Systemorganklasse; UE: Unerwünschtes Ereignis.

Für die Studie RISE wurden nur UE nach SOC und PT dargestellt, die bei mindestens 2 Personen aufgetreten sind. Nur die SOC „Erkrankungen des Gastrointestinaltrakts“ und „Infektionen und parasitäre Erkrankungen“ erfüllten dieses Kriterium.

*Tabelle 22: UE mit Inzidenz ≥ 10 % und aufgetreten bei mind. 2 Personen; Studie RISE, Sicherheitspopulation<sup>1)</sup>*

Studie RISE UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	MRX N = 4 n (%)
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>2 (50,0)</b>
<b>Infektionen und parasitäre Erkrankungen</b>	<b>3 (75,0)</b>
Coronavirus-Infektion	2 (50,0)

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

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; MRX: Maralixibat; PT: Preferred Term; SOC: Systemorganklasse; UE: Unerwünschtes Ereignis.

### Schwere unerwünschte Ereignisse

Sowohl in der Gesamtpopulation als auch der PFIC-Population der Studie MARCH-PFIC trat kein schweres UE nach SOC und PT bei  $\geq 5$  % der Personen in mindestens einem Studienarm auf.

In der Studie RISE sind keine schweren UE dokumentiert.

### Schwerwiegende unerwünschte Ereignisse

Sowohl in der Gesamtpopulation als auch der PFIC-Population der Studie MARCH-PFIC trat kein SUE nach SOC und PT bei  $\geq 5$  % der Personen in mindestens einem Studienarm auf.

In der Studie RISE ereignete sich ein SUE (25 %) des PT „Gastroenteritis durch Adenovirus“ in der SOC „Infektionen und parasitäre Erkrankungen“.

### Unerwünschte Ereignisse von besonderem Interesse

UE von besonderem Interesse für die Studie MARCH-PFIC sind in Tabelle 23 aufgeführt. In der PFIC-Population traten lediglich 2 schwerwiegende UE von besonderem Interesse auf, ein Ereignis in der Maralixibat-Gruppe (erhöhtes Bilirubin), ein Ereignis in der Placebo-Gruppe (Mangel an fettlöslichen Vitaminen). Kein UE von besonderem Interesse ist als „schwer“ angegeben. Einzig das UE von besonderem Interesse „Diarrhö“ trat statistisch signifikant häufiger unter Maralixibat auf.

Für die Studie RISE ist kein UE von besonderem Interesse dokumentiert.

Tabelle 23: UE von besonderem Interesse; Studie MARCH-PFIC, Sicherheitspopulation

Studie MARCH-PFIC UE von besonderem Interesse <sup>1)</sup> Jeglicher Schweregrad	PFIC-Population		RR [95%-KI]; p-Wert	Gesamtpopulation		RR [95%-KI]; p-Wert
	MRX N = 33 n (%)	PBO N = 31 n (%)		MRX N = 47 n (%)	PBO N = 46 n (%)	
Diarrhö <sup>2)</sup>	20 (60,6)	5 (16,1)	3,76 [1,61; 8,78]; 0,0003	29 (61,7)	9 (19,6)	k. A.
Mangel an fettlöslichen Vitaminen <sup>3)</sup>	7 (21,2)	10 (32,3)	0,66 [0,29; 1,51]; 0,321	12 (25,5)	15 (32,6)	k. A.
Erhöhte Transaminasen <sup>4)</sup>	4 (12,1)	3 (9,7)	1,25 [0,30; 5,15]; 0,756	7 (14,9)	3 (6,5)	k. A.
Erhöhtes Bilirubin <sup>5)</sup>	5 (15,2)	7 (22,6)	0,67 [0,24; 1,89]; 0,450	7 (14,9)	9 (19,6)	k. A.

<sup>1)</sup> In den Studienunterlagen wird der Begriff „Ereignis von klinischem Interesse“ definiert, im Dossier dagegen als „UE von besonderem Interesse“ bezeichnet.

<sup>2)</sup> Umfasst UE der PT und LLT „Diarrhö“, „Intermittierende Diarrhö“, „Erhöhte Stuhlentleerung“ und „Gastroenteritis“.

<sup>3)</sup> Umfasst viele im Anhang des SAP gelistete, SOC-übergreifende PT.

<sup>4)</sup> Umfasst UE der PT und LLT „Erhöhte Leberenzyme“, „Transaminasen anomal“, „Transaminasen erhöht“, „Alanin-Aminotransferase erhöht“, „Aspartat-Aminotransferase erhöht“, „Aspartat-Aminotransferase anomal“ und „Alanin-Aminotransferase anomal“.

<sup>5)</sup> Umfasst UE der PT und LLT „Bilirubin im Blut erhöht“, „Bilirubin im Blut anomal“ und „Hyperbilirubinämie“.

Abkürzungen: k. A.: keine Angabe; KI: Konfidenzintervall; LLT: Lowest Level Term; MRX: Maralixibat; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; PT: Preferred Term; RR: Relatives Risiko; 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 der progressiven familiären intrahepatischen Cholestase (PFIC) bei Patientinnen und Patienten ab dem Alter von 3 Monaten. 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. Es handelt sich um eine Zulassungserweiterung. Die initiale Zulassung bezieht sich auf die Behandlung des cholestatischen Pruritus bei Patientinnen und Patienten mit Alagille-Syndrom (ALGS) ab dem Alter von 2 Monaten. Die empfohlene Zieldosis (Erhaltungsdosis) im Anwendungsgebiet PFIC liegt bei 570 µg/kg BID. Die Anfangsdosis beträgt 285 µg/kg QD und kann nach ein bis zwei Wochen auf 285 µg/kg BID und nach weiteren ein bis zwei Wochen auf 570 µg/kg BID erhöht werden, falls klinisch angezeigt und je nach Verträglichkeit. Maralixibat 570 µg/kg als freie Base entspricht einer Dosis von Maralixibat-Chlorid von 600 µg/kg. Weiterhin liegt ein allgemeiner Warnhinweis für Kinder unter 5 Jahren vor, der die empfohlene Höchstdosis von Maralixibat aufgrund des Propylenglykol-Gehalts auf 285 µg/kg BID begrenzt.

Die Zulassung basiert auf der pivotalen Studie MARCH-PFIC, in die Kinder und Jugendliche von 1 Jahr bis ausschließlich 18 Jahren mit PFIC, einer bestehenden Cholestase und Pruritus eingeschlossen waren. Zusätzlich wurde Evidenz aus der Studie RISE herangezogen, die Säuglinge (< 12 Monate) mit cholestatischen Lebererkrankungen (PFIC oder ALGS) einschloss. Beide Studien wurden in der vorliegenden Nutzenbewertung berücksichtigt.

Im Vergleich zur Fachinformation ergeben sich bei der Applikation folgende Abweichungen für die Studien MARCH-PFIC und RISE:

- Dosisescalation auf 600 µg/kg BID regelhaft für alle Patientinnen und Patienten vorgesehen, wohingegen die Fachinformation eine Dosiserhöhung auf die maximale Erhaltungsdosis nur vorsieht, falls klinisch angezeigt und je nach Verträglichkeit.
- Keine Beschränkung der Maximaldosis auf 300 µg/kg BID für Kinder bis 5 Jahre.
- Längere Auftitrationsperiode mit mehr Titrationschritten (2 Titrationschritte in 2–4 Wochen vs. 4 bzw. 3 Titrationschritte in 4–6 Wochen in den Studien MARCH-PFIC und RISE).

Die Auswirkungen der abweichenden Auftitrationsperiode werden als gering eingestuft. Sowohl die Abweichung in der Vorgabe zur Dosisescalation als auch der Beschränkung der Maximaldosis für Kinder bis 5 Jahre führt jedoch regelhaft zu deutlich niedrigeren Dosierungen in der Versorgung verglichen mit den zulassungsbegründenden Studien. Während Maralixibat auch in der hohen Dosierung von 600 µg/kg BID insgesamt gut verträglich war (siehe Kapitel 3.5), ist die Auswirkung dieser Abweichungen auf die Wirksamkeit, insbesondere bei den Kindern bis 5 Jahren, nicht zu beurteilen. Gemäß EPAR zeigen sich Verbesserungen in Pruritus und sBA bereits zeitnah nach Beginn der Dosisescalationsperiode bei deutlich geringeren Dosierungen von Maralixibat. Nach weiterer Eskalation auf 600 µg/kg BID wird jedoch ein zusätzliches Ansprechen auf die Therapie beobachtet. In Bezug auf die Nutzenbewertung können die Ergebnisse der klinischen Studien nicht direkt auf die Population der Kinder bis 5 Jahre übertragen werden.

Auch hinsichtlich der Population bestehen Abweichungen zwischen den eingeschlossenen Personen der Studien MARCH-PFIC und RISE und dem zugelassenen Anwendungsgebiet:

- Keine klinischen Daten für Erwachsene (Personen ab 18 Jahren); der Anteil der Jugendlichen ab einschließlich 13 Jahren in der Studie MARCH-PFIC war mit 6 % gering.
- Keine klinischen Daten für Personen mit PFIC5 und nur geringe Anteile < 10 % von Personen mit PFIC3, PFIC4 und PFIC6 an der Studienpopulation von MARCH-PFIC.
- Keine kontrollierten Daten für Kinder zwischen 3 und 12 Monaten.

- Ein cholestatischer Pruritus (Einschlusskriterium für die Studien MARCH-PFIC und RISE) ist gemäß Zulassungstext nicht erforderlich.
- Keine Ergebnisse hinsichtlich der Wirksamkeit für Personen mit t-PFIC2, Heterozygotie, niedrigem oder fluktuierendem sBA oder früherer Operation zur Behandlung der PFIC (siehe auch Kapitel 4.2).

In Bezug auf den letzten Punkt weist die Fachinformation einschränkend darauf hin, dass Erkrankungen, Arzneimittel oder chirurgische Eingriffe, die entweder die gastrointestinale Motilität oder die enterohepatische Zirkulation von Gallensäuren (einschließlich des Gallensalztransports in die kleinsten Gallengänge) beeinträchtigen, die Wirksamkeit von Maralixibat vermindern können. Aus diesem Grund sei bei Personen mit PFIC2, bei denen das Protein BSEP vollständig fehlt oder nicht funktioniert (d. h. Personen mit t-PFIC2), kein Ansprechen auf Maralixibat zu erwarten. Ähnliches gilt für Kinder, bei denen keine etablierte genetische Variante im Zusammenhang mit PFIC gefunden wurde oder für Patientinnen und Patienten mit vorliegender Heterozygotie, bei denen nur ein Allel des betroffenen Gens mutiert, während das andere Allel noch funktionsfähig ist. Deren Erkrankung kann sich anders bzw. schwächer präsentieren. Eine Behandlung mit Maralixibat ist bei diesem Personenkreis jedoch nicht ausgeschlossen. Im EPAR wird hierzu angeführt, dass die Interpretation genetischer Testergebnisse keine hundertprozentig sichere Prognose über das vollständige Fehlen des BSEP-Proteins (t-PFIC2) erlaubt, weshalb eine residuale BSEP-Aktivität nicht ausgeschlossen werden kann. Der Anteil der Personen mit t-PFIC2 in der Studie MARCH-PFIC war mit 10,6 % im Interventionsarm bzw. 8,7 % im Kontrollarm zudem relativ gering.

Die beschriebenen Abweichungen führen dazu, dass für einen Teil des Anwendungsgebiets keine oder nur wenige klinische Daten bzw. Wirksamkeitsdaten vorliegen. Die Datenlage zur Prävalenz der Subtypen in der Allgemeinbevölkerung ist limitiert [2,4]. Die Subtypen PFIC1, PFIC2 und PFIC3 sind am besten beschrieben und treten am häufigsten auf. Zum Anteil der Personen mit anderen Subtypen oder „keinem etablierten Genotyp“ konnten keine Informationen identifiziert werden. Im EPAR wird von einer Übertragbarkeit der Studienergebnisse auf Personen mit PFIC5 ausgegangen. Insgesamt verbleiben jedoch Unsicherheiten, was die Übertragbarkeit der Wirksamkeits- und Sicherheitsergebnisse von Maralixibat in den Studien MARCH-PFIC und RISE auf alle vom Anwendungsgebiet umfassten Patientinnen und Patienten betrifft.

Vor dem Hintergrund der placebokontrollierten Studie MARCH-PFIC ist hinsichtlich der Übertragbarkeit auf den deutschen Versorgungskontext darauf hinzuweisen, dass mit Odevixibat im vorliegenden Anwendungsgebiet PFIC ab einem Alter von 6 Monaten seit kurzem bereits eine Therapiealternative mit geringem Zusatznutzen in Deutschland zugelassen ist.

Die Zulassungserweiterung von Maralixibat erfolgte unter „außergewöhnlichen Umständen“.

## 4.2 Design und Methodik der Studie

### Studie MARCH-PFIC

Die Nutzenbewertung stützt sich insbesondere auf die pivotale Studie MARCH-PFIC, einer multizentrischen, doppelblinden, placebokontrollierten Studie zur Untersuchung der Wirksamkeit und Sicherheit von Maralixibat bei Kindern und Jugendlichen mit PFIC. Auf eine bis zu 6-wöchige Screeningperiode folgt eine offene Dosisescalationsphase über 4–6 Wochen, in der Maralixibat auf bis zu 600 µg/kg BID auftitriert wurde und eine 20- bis 22-wöchige Behandlungsphase mit stabiler Dosierung. Nach Woche 26 bestand die Möglichkeit einer Teilnahme an der einarmigen Extensionsstudie MRX-503. Für Personen, die nicht an der Extensionsstudie teilnehmen, erfolgte eine Sicherheitsnachbeobachtung 7 Tage nach Einnahme der letzten Dosis der Studienmedikation sowie jährliche Follow-up-Visiten. Dahingehende Daten konnten den Unterlagen nicht entnommen werden. In die Studie MARCH-PFIC wurden 93 Personen aus 29 Studienzentren in 16 Ländern

weltweit eingeschlossen, darunter auch Zentren aus Deutschland. Die Studie wurde von der letzten Person am 01.09.2022 abgeschlossen und die Ergebnisse basieren auf dem finalen Datenschnitt (14.10.2022). Primärer Endpunkt war die Veränderung der Pruritus-Schwere gemessen mittels ItchRO(Obs) von Baseline zu Woche 15–26. Weitere Endpunkte waren u. a. „sBA“, „Klinische Kratzskala“, „Körperliche Entwicklung“, „Schlafstörungen mittels EDQ(Obs)“, Fatigue und Lebensqualität mittels „PedsQL“ sowie „Unerwünschte Ereignisse“.

Begleitmedikationen waren erlaubt, sofern sie ab 30 Tage vor der Screening-Visite stabil gehalten wurden. Der Bedarf einer PEBD in Folge einer Krankheitsprogression oder Lebertransplantation (Aufnahme auf Transplantationsliste) sind u. a. als Gründe für einen Studienabbruch beschrieben.

Beim Studieneinschluss wurde zwischen 2 Kohorten anhand der Ein- und Ausschlusskriterien unterschieden. Die primäre Kohorte, randomisiert 1:1 auf Maralixibat oder Placebo, bestehend aus Personen mit nt-PFIC2, ausgenommen von Personen mit Heterozygotie, mit niedrigem oder fluktuierendem sBA oder mit früherer Operation zur Behandlung der PFIC. Und die „ergänzende Kohorte“, randomisiert im Verhältnis 1:1 auf Maralixibat oder Placebo, stratifiziert nach PFIC1, PFIC3 und allen weiteren Studienteilnehmenden der Kohorte; darin wurden alle Studienteilnehmenden aufgenommen, auf die die Kriterien der primären Kohorte nicht zutrafen, also z. B. Personen mit t-PFIC2, anderen PFIC-Varianten oder intermittierender Cholestase. Aus diesen Kohorten wurden wiederum 3 unterschiedliche, sich teilweise überschneidende Populationen für die Auswertung gebildet:

- Primäre-Population; alle Personen der primären Kohorte.
- PFIC-Population; primäre Kohorte sowie alle bekannten etablierten Genotypen, ausgenommen Personen mit t-PFIC2, Heterozygotie, niedrigem oder fluktuierendem sBA oder früherer Operation zur Behandlung der PFIC.
- Gesamtpopulation; alle Studienteilnehmenden.

Da sich die Primäre-Population ausschließlich auf nt-PFIC2 beschränkt und damit nur einen Teil des zugrundeliegenden Anwendungsgebiets darstellt, wird sie für die Nutzenbewertung nicht berücksichtigt. Als maßgeblich für die vorliegende Nutzenbewertung wird die Gesamtpopulation erachtet, da aus der Fachinformation kein expliziter Behandlungsausschluss bestimmter Patientengruppen hervorgeht (siehe auch Kapitel 4.1). Dies steht im Gegensatz zum Vorgehen des pU im Nutzen-dossier, der die PFIC-Population und die Primäre-Population berücksichtigt, da alle darüber hinaus eingeschlossenen Patientinnen und Patienten nicht vom Anwendungsgebiet umfasst seien. Da keine Ergebnisdaten zur Wirksamkeit in der Gesamtpopulation vorliegen, werden ergänzend die Ergebnisse der PFIC-Population dargestellt.

Für den Einschluss in die Studie MARCH-PFIC wurden 125 Personen gescreent, wovon letztendlich 93 eingeschlossen und behandelt wurden. Die ITT-Population entspricht der Sicherheitspopulation. Der Anteil der Personen mit Abbruch der Studienmedikation und damit auch Studienabbruch war insgesamt gering, mit 3 Personen (6 %) der Maralixibat-Gruppe und 4 (9 %) der Placebo-Gruppe. 77 % der Personen aus dem Interventionsarm und 83 % aus dem Placebo-Arm gingen nach Studienende in die Extensionsstudie MRX-503 über. In der Gesamtpopulation wurden 45 Personen (96 %) des Maralixibat-Arms auf die Maximaldosis von 600 µg/kg BID hochtitriert. Die mediane Behandlungsdauer für den Maralixibat-Arm beträgt in beiden Populationen 183 Tage, für den Placebo-Arm liegen keine Angaben vor.

In der Gesamtpopulation der Studie befanden sich 5 Geschwisterpaare mit insgesamt 10 Geschwisterkindern (11 %), wobei beide Kinder eines Geschwisterpaares verblindet in dieselbe Behandlungsgruppe randomisiert wurden. Die Kinder und Jugendlichen waren im Median mit 3 Jahren (Interventionsgruppe) bzw. 3,5 Jahren (Placebo-Gruppe) sehr jung, nur 4 (9 %) bzw. 2 Personen (4 %) waren 13 Jahre oder älter. Sowohl bei der Körpergröße als auch beim Körpergewicht zeigten sich deutlich negative Abweichungen von den Normwerten. Zu Baseline nahmen bereits über 80 % der Patientinnen und Patienten UDCA ein sowie ca. 50 % Rifampicin. Hinsichtlich des PFIC-Genotyps

wurde bei knapp 40 % der Gesamtpopulation nt-PFIC2 diagnostiziert und bei weiteren ca. 10 % t-PFIC2; PFIC-1 machte knapp 20 % der Gesamtpopulation aus und die weiteren Genotypen jeweils unter 10 % der Gesamtpopulation. „Keine etablierte Variante gefunden“ traf nur auf 3 Personen (6 %) im Maralixibat-Arm und 5 (11 %) im Placebo-Arm zu. Lediglich 2 Personen (4 %) des Placebo-Arms wiesen eine heterozygote Variante auf. Aus der Gesamtpopulation hatten 8 Personen (9 %) bereits eine PFIC-assoziierte Operation erhalten (5 im Maralixibat-Arm, 3 im Placebo-Arm) und jeweils 1 Person aus jedem Behandlungsarm wies fluktuierende sBA-Werte auf.

Insgesamt waren die Baseline-Charakteristika auf Ebene der Gesamtpopulation zwischen den Studienarmen ausgeglichen. Innerhalb der PFIC-Population zeigten sich zu Baseline Unterschiede zwischen Interventions- und Kontrollgruppe bei dem Einsatz von UDCA (82 vs. 97 %), dem z-Score für Körpergewicht (Median: -1,46 vs. -1,09), der Krankheitsdauer in Monaten (Median: 28 vs. 12) und dem Anteil von PFIC4 (18 vs. 3 %). Inwiefern dies durch die kleinere Stichprobe oder die zum Teil fehlende Stratifizierung in der PFIC-Population begründet ist, kann nicht abschließend beurteilt werden.

Innerhalb der Studienunterlagen von MARCH-PFIC fallen geringe Diskrepanzen in den Ergebnisdaten zwischen Studienbericht und den statistischen Outputs auf, z. B. bei den stetigen Analysen zum ItchRO(Obs) oder den Ergebnissen zum sBA. Gründe hierfür konnten nicht identifiziert werden. Die Ergebnisse sind jedoch in jedem Fall gleichgerichtet.

Das Verzerrungspotential auf Studienebene für die Gesamtpopulation wird als gering eingestuft. Für die Teilpopulation „PFIC-Population“ wird von einem unklaren Verzerrungspotential ausgegangen.

### **Studie RISE**

Bei der ergänzend im EPAR berücksichtigten Studie RISE handelt es sich um eine offene, multi-zentrische 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 PFIC relevant. Nach einer 4-wöchigen Screeningphase wurde Maralixibat in 4 bis 6 Wochen auf eine Dosis von maximal 600 µg/kg BID oder die höchste tolerierte Dosis auftitriert, die in der anschließenden Erhaltungsphase von Woche 5 oder 7 bis Woche 13 weiter verabreicht wurde. Eine Weiterbehandlung in der Langzeitphase 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“ und „Leberassoziierte Ereignisse“ auch „ItchRO“, „Klinische Kratzskala“ und „Körperliche Entwicklung“ als Endpunkte definiert. Abweichend von der Studie MARCH-PFIC war das Vorhandensein eines cholestatischen Pruritus kein explizites Einschlusskriterium.

Bis zum Interim-Datenschnitt vom 23.06.2022 wurden 4 Säuglinge eingeschlossen und mit Maralixibat behandelt. Alle haben bereits die 13-wöchige Kernstudie beendet und befanden sich zum Datenschnitt in der langfristigen Behandlungsphase ab Woche 13. Die mediane Behandlungsdauer liegt bei 115 Tagen. Das mediane Alter der Patientinnen und Patienten zu Baseline betrug 8,5 Monate und jeweils 2 Säuglinge waren männlich bzw. weiblich. Es zeigen sich bereits deutlich negative Abweichungen bei den Parametern der körperlichen Entwicklung im Vergleich zur Norm. Bei 2 Säuglingen wurde PFIC2 diagnostiziert, bei jeweils einem weiteren PFIC1 und PFIC4. Auf Basis der klinischen Kratzskala wird deutlich, dass alle 4 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. Im EPAR werden

Ergebnisse aus einem aktualisierten Datenschnitt der Studie RISE vom 28.06.2023 dargestellt, den der pU nicht mit eingereicht und in das Nutzendossier einbezogen hat. Zu diesem Zeitpunkt waren bereits 10 Personen mit PFIC in die Studie eingeschlossen. Es ist unklar, ob dieser weitere Interim-Datenschnitt von der Zulassungsbehörde gefordert wurde.

Abgesehen davon erscheint es allgemein fraglich, wie valide die PFIC-Symptomatik gerade in den ersten Lebensmonaten erfasst werden kann bzw. inwiefern 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 nicht abschließend bewertet werden kann.

Da es sich bei der Studie RISE um eine einarmige Studie handelt, wird von einem hohen Verzerrungspotential ausgegangen.

### **4.3 Mortalität**

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

Eine Interpretation und Bewertung der Daten ist aufgrund der relativ kurzen Studiendauer von MARCH-PFIC nicht möglich und der Effekt von Maralixibat auf die Mortalität in der Indikation PFIC kann auf Grundlage der vorgelegten Daten nicht beurteilt werden.

### **4.4 Morbidität**

Für die Morbidität legt der pU ausschließlich Ergebnisse zur Wirksamkeit in der PFIC-Population der Studie MARCH-PFIC vor, die ergänzend dargestellt werden. Aus Transparenzgründen wurden Ergebnisse zum sBA-Spiegel im Anhang berichtet. Der Endpunkt „Klinische Kratzskala“ wird als weitere Doppelerfassung des Pruritus von unklarer Validität lediglich ergänzend zum ItchRO dargestellt. Auch die Darstellung zu „Leberassoziierte Ereignisse“ erfolgte lediglich ergänzend.

#### **Pruritus mittels ItchRO**

Für den ItchRO liegen eine Vielzahl an Auswertungen für die PFIC-Population in der Studie MARCH-PFIC vor. Dabei werden primär die Responderanalysen als relevant erachtet. Diese liegen jedoch ausschließlich für die Erhebungen am Morgen vor. Zudem weisen die Responderanalysen teilweise Mängel auf, die die Herleitung und Begründung der Responderkriterien betreffen. Daher werden zusätzlich Gruppenunterschiede in der Veränderung der Mittelwerte (morgendlich, abendlich, höchster täglicher Wert) berichtet. Die Analysen zur Häufigkeit des Pruritus gemessen anhand des ItchRO werden nicht dargestellt, da die Häufigkeit als ein Teilaspekt der Schwere angesehen wird.

Angaben zur Anzahl fehlender Werte bzw. dem Anteil der fehlenden Tagebucheinträge zur Berechnung der Rücklaufquote der einzelnen Auswertungen konnten den Studienunterlagen weder für den ItchRO(Obs) noch für den ItchRO(Pt) entnommen werden. Im Rahmen einer Post-hoc-Analyse im Nutzendossier gibt der pU an, dass die Patientinnen und Patienten der Interventionsgruppe kumuliert insgesamt 6.126 Tage im gesamten Beobachtungszeitraum waren, davon ist an 5.562 Tagen (90,8 %) eine Erhebung des ItchRO dokumentiert. Für die Kontrollgruppe werden 5.311 Tage im Beobachtungszeitraum berichtet, mit insgesamt 4.970 Tagen (93,6 %) mit ItchRO-Erhebungen. Obwohl eine direkte Berechnung spezifischer Rücklaufquoten für einzelne Analysen aus diesen Angaben nicht möglich ist, erscheint die Datenvollständigkeit des Endpunkts „ItchRO“ auch unter Berücksichtigung der Studienabbrüche in einem Maße gegeben, sodass die Ergebnisse insgesamt berücksichtigt werden können.

In der berücksichtigten präspezifizierten Responderanalyse war der Anteil der Personen mit einer Verbesserung  $\geq 1$  Punkt oder mit einem Schweregrad-Score  $\leq 1$  zu Woche 15–26 (RR: 2,47 (95%-KI: [1,29; 4,72])) unter Maralixibat statistisch signifikant höher im Vergleich zu Placebo.

In den ergänzend berücksichtigten kontinuierlichen Auswertungen des ItchRO(Obs) ergab sich ebenfalls ein statistisch signifikanter Unterschied zugunsten von Maralixibat in der Veränderung des mittleren morgendlichen ItchRO(Obs)-Schweregrad-Scores in Woche 15–26 gegenüber Baseline (LS-MWD: -1,20 (95%-KI: [-1,73; -0,67])), der gemessen am Hedges' g auch klinisch relevant ist. Neben dieser als primär definierten Analyse zeigten die Veränderung des abendlichen Schweregrad-Scores und dem Maximum aus Abend- und Morgenerhebung für denselben Zeitraum Effekte von vergleichbarer Stärke und Signifikanz. Auch die durchgeführten Sensitivitätsanalysen sowie die Betrachtung abweichender Studienzeiträume zeigten gleichgerichtete Ergebnisse.

Der ItchRO(Pt) wurde nur bei einer kleinen Subpopulation der  $\geq 9$ -Jährigen erhoben. Zu Baseline lagen lediglich Daten für 5 Teilnehmende aus der Maralixibat-Gruppe und für 4 Teilnehmende aus der Placebo-Gruppe vor. Die Ergebnisse für den ItchRO(Pt) wurden in der Nutzenbewertung nicht berichtet.

Zusammenfassend zeigen sich für den ItchRO(Obs) sowohl in der Responderanalyse als auch den kontinuierlichen Auswertungen konsistent statistisch signifikante Unterschiede zugunsten von Maralixibat. Insgesamt wird deshalb von einem Vorteil von Maralixibat beim Pruritus gegenüber Placebo ausgegangen, auch wenn keine selbstberichteten Daten aus dem ItchRO(Pt) berücksichtigt werden konnten. Das Verzerrungspotential des ItchRO(Obs) wird mangels fehlender Angaben zu den Rücklaufquoten für die spezifischen Auswertungen mit unklar bewertet.

Für die Gesamtpopulation der Studie MARCH-PFIC und die Studie RISE liegen keine Ergebnisse zum Endpunkt vor.

### **Schlafstörungen mittels EDQ(Obs)**

Zur Beurteilung Pruritus-bedingter Schlafstörungen wurde in der Studie MARCH-PFIC die Schlafbezogene Frage aus dem EDQ(Obs) am Morgen ausgewertet. Unklar ist, warum Schlafstörungen lediglich im betreuerberichteten Tagebuch EDQ(Obs) erhoben wurden und nicht auch im patientenberichteten EDQ(Pt). Da eine Fremdbewertung des EDQ(Obs) nur bei Kindern unter 9 Jahren vorgesehen war, liegen Ergebnisse nur für eine Teilpopulation vor, deren genaue Größe unklar ist. Zu Baseline lagen Angaben von 27 Kindern aus dem Maralixibat-Arm und 26 aus dem Placebo-Arm vor. Angaben zu Rücklaufquoten konnten nicht identifiziert werden. Die Ergebnisse zu den Pruritus-bedingten Schlafstörungen werden nicht berücksichtigt.

Für die Gesamtpopulation der Studie MARCH-PFIC sowie die Studie RISE liegen keine Ergebnisse zum Endpunkt vor.

### **Körperliche Entwicklung**

Die körperliche Entwicklung wurde anhand der Veränderungen in den anthropometrischen Parametern „Körpergröße“ und „Körpergewicht“ unter Berücksichtigung standardisierter z-Scores in der Studie MARCH-PFIC sehr standardisiert erhoben und evaluiert.

In der PFIC-Population zeigte sich in der Auswertung der Wochen 18–26 gegenüber Baseline in der Körpergröße kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen. Bei deutlich höheren negativen Abweichungen im Körpergewicht (z-Score) zu Baseline im Maralixibat-Arm (-1,75) im Vergleich zum Placebo-Arm (-1,28), zeigt sich ein statistisch signifikanter Unterschied in der Veränderung zu Woche 18–26 zugunsten von Maralixibat (LS-MWD: 0,23 (95%-KI: [0,01; 0,44])) mit einem Hedges' g von 0,53 (95%-KI: [0,02; 1,03]).

Das Verzerrungspotential des Endpunkts „Körperliche Entwicklung“ wird aufgrund der bestehenden Unterschiede im Körpergewicht zu Baseline in der PFIC-Population mit hoch bewertet. Es ist zudem unklar, ob die Studienlänge von MARCH-PFIC ausreichend ist, um die Effekte von Maralixibat auf die körperliche Entwicklung erfassen und bewerten zu können.

Für die Gesamtpopulation der Studie MARCH-PFIC und die Studie RISE liegen keine Ergebnisse zum Endpunkt vor.

### **PedsQL-Fatigue**

Die PedsQL-Fatigue wurde in der Studie MARCH-PFIC in den unterschiedlichen altersspezifischen Versionen im Elternbericht ab 2 Jahren und im Selbstbericht ab 8 Jahren ausgefüllt. Wie viele Kinder in der Studie insgesamt 2 bzw. 8 Jahre oder älter sind, ist den Studienunterlagen nicht zu entnehmen. Eine Auswertung der selbstberichteten Versionen war weder geplant, noch wurden Ergebnisse im Nutzendossier dargestellt. Eine Rationale für dieses Vorgehen ist nicht ersichtlich. Angaben zu Rücklaufquoten liegen ebenfalls nicht vor, sodass die Ergebnisse der PedsQL-Fatigue nicht berücksichtigt werden.

Für die Gesamtpopulation der Studie MARCH-PFIC und die Studie RISE liegen keine Ergebnisse zum Endpunkt vor.

## **4.5 Lebensqualität**

Analog zur PedsQL-Fatigue wurden auch beim Kernmodul des PedsQL in der PFIC-Population der Studie MARCH-PFIC keine Responderanalysen präsentiert, die Daten der von den Kindern selbst eingeschätzten Version wurden nicht ausgewertet und es fehlen Informationen zu den Rücklaufquoten. Die Ergebnisse des Endpunkts werden nicht für die Nutzenbewertung herangezogen.

Für die Gesamtpopulation der Studie MARCH-PFIC und die Studie RISE liegen keine Ergebnisse zum Endpunkt vor.

## **4.6 Sicherheit**

Ergebnisse für den Anteil der Personen mit UE liegen in der Studie MARCH-PFIC sowohl für die Gesamtpopulation als auch für die PFIC-Population vor. Statistisch post hoc ausgewertet wurden jedoch lediglich die UE der PFIC-Population und hinsichtlich der Spezifikation des statistischen Tests bestehen Unklarheiten. Eine Erfassung der UE war lediglich bis 7 Tage nach Verabreichung der letzten Dosis der Studienmedikation vorgesehen. Sofern ein Übergang in die Extensionsstudie erfolgte, wurde nicht weiter nachbeobachtet. Hinsichtlich der Beobachtungsdauer liegen keine Daten für die Behandlungsarme vor. Ergebnisse zur Behandlungsdauer existieren nur für den Interventionsarm. Aufgrund der geringen Anzahl an Studienabbrüchen und der Angaben zur Datenvollständigkeit der Tagebucheinträge beim ItchRO ist jedoch von einer vergleichbaren Beobachtungsdauer der Studienarme auszugehen.

Bei den Gesamtraten zeigten sich in beiden Populationen keine wesentlichen Unterschiede zwischen den Behandlungsgruppen. Von schweren UE und SUE waren in beiden Populationen nur wenige Patientinnen und Patienten betroffen (Anteil jeweils unter 10 %) und die Unterschiede in der PFIC-Population waren statistisch nicht signifikant zwischen den Behandlungsgruppen. Der einzig dokumentierte Studienabbruch aufgrund eines UE ereignete sich in der Maralixibat-Gruppe der Gesamtpopulation.

Die häufigsten UE in beiden Populationen waren die PT „Diarrhö“ und „Fieber“ aus den SOC „Erkrankungen des Gastrointestinaltrakts“ und „Infektionen und parasitäre Erkrankungen“. In der PFIC-Kohorte waren die Unterschiede für „Diarrhö“ (RR: 3,57 (95%-KI: [1,52; 8,39])) und „Erkrankungen

des Gastrointestinaltrakts“ (RR: 1,96 (95%-KI: [1,21; 3,17])) jeweils statistisch signifikant zuungunsten von Maralixibat. Weiterhin zeigte sich ein statistisch signifikanter Effekt bei der SOC „Leber- und Gallenerkrankungen“ (RR: 0,13 (95%-KI: [0,02; 1,03])) zuungunsten von Placebo. Für die Gesamtpopulation liegen keine Effektschätzer vor, es ist mit ähnlichen Ergebnissen zu rechnen. Sowohl bei den schweren UE als auch den SUE ereignete sich kein UE einer SOC oder eines PT bei mindestens 5 % der Patientinnen und Patienten eines Studienarms.

In Anbetracht der häufiger in der Placebo-Gruppe aufgetretenen SOC und PT „Leber- und Gallenerkrankungen“, „Hyperbilirubinämie“ und „Pruritus“ deutet sich eine Erfassung der Symptomatik der Grunderkrankung in den UE an, obwohl Symptome nicht als UE erfasst werden sollten, solange sie innerhalb der normalen täglichen Schwankung oder der erwarteten Progression der Erkrankung liegen und Teil der in der Studie zu erfassenden Wirksamkeitsdaten sind. Auswertungen unter Nicht-Berücksichtigung erkrankungsbezogener Ereignisse liegen nicht vor.

UE von besonderem Interesse waren gemäß Studienunterlagen nicht definiert. Im Nutzendossier des pU wurden die a priori definierten „Ereignisse von klinischem Interesse“ als „UE von besonderem Interesse“ bezeichnet. Einschränkend ist darauf hinzuweisen, dass die Ereignisse „Erhöhtes Bilirubin“, „Erhöhte Transaminasen“ und „Mangel an fettlöslichen Vitaminen“ inhärente Merkmale der Grunderkrankung PFIC darstellen. Lediglich das UE von besonderem Interesse „Diarrhö“ ist als spezifische Nebenwirkung von Maralixibat bekannt. Ausschließlich bei „Diarrhö“ jeglichen Schweregrads zeigte sich auch ein statistisch signifikanter Unterschied zuungunsten von Maralixibat in der PFIC-Population. Da die Ereignisraten auf Ebene der Gesamtpopulation mit 62 % unter Maralixibat und 20 % unter Placebo ähnlich sind, ist auch hier von einem statistisch signifikanten Effekt zuungunsten von Maralixibat auszugehen.

In der Studie RISE ereignete sich bei allen Patientinnen und Patienten mindestens ein UE und 1 Kleinkind (25 %) erlitt ein SUE des PT „Gastroenteritis durch Adenovirus“ in der SOC „Infektionen und parasitäre Erkrankungen“. Schwere UE und UE, die zum Therapieabbruch führen, traten nicht auf. UE von besonderem Interesse ereigneten sich nicht. Lediglich UE der SOC „Infektionen und parasitäre Erkrankungen“ und „Erkrankungen des Gastrointestinaltrakts“ sowie des PT „Coronavirus-Infektion“ ereigneten sich bei 2 oder mehr Kleinkindern.

Das Verzerrungspotential der UE aus der Studie MARCH-PFIC für die PFIC-Population wird aufgrund des unklaren Verzerrungspotentials auf Studienebene als unklar beurteilt. Vor dem Hintergrund des geringen Stichprobenumfangs und der Einschränkungen bei der Erhebung und Auswertung der Sicherheitsdaten hinsichtlich der Erfassung von Krankheitssymptomatik und der kurzen Nachbeobachtungsdauer, ist eine abschließende Bewertung der Sicherheit von Maralixibat bei Personen mit PFIC nicht möglich.

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

Die Angaben der Fachinformation sind zu berücksichtigen.

## 6 Zusammenfassung der Nutzenbewertung

Maralixibat (Livmarli®) ist zugelassen für die Behandlung der progressiven familiären intrahepatischen Cholestase (PFIC) bei Patienten ab dem Alter von 3 Monaten. Die Nutzenbewertung von Maralixibat basiert auf der pivotalen Studie MARCH-PFIC sowie der supportiven Studie RISE. Die Studie MARCH-PFIC ist eine 26-wöchige randomisierte, placebokontrollierte und doppelblinde Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit von Maralixibat bei Patientinnen und Patienten mit PFIC ab einem Alter von 12 Monaten bis 18 Jahren. Die Studie RISE ist eine offene Phase-II-Studie zur Untersuchung der Sicherheit und Verträglichkeit von Maralixibat bei Säuglingen (< 12 Monate).

Die Ergebnisse der in der Nutzenbewertung berücksichtigten Endpunkte der Studien werden in der folgenden Tabelle 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 24: Zusammenfassende Darstellung der Ergebnisse der Studie MARCH-PFIC

Studie MARCH-PFIC Darstellung der Ergebnisse	Maralixibat			Placebo			Maralixibat vs. Placebo	Effekt
<b>Mortalität</b>								
<b>Todesfälle</b>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>		<i>N</i>	<i>Personen mit Ereignis, n (%)</i>		<i>RR [95%-KI]; p-Wert</i>	
PFIC-Population	33	0		31	0		-	↔
Gesamtpopulation	47	0		46	0		-	↔
<b>Morbidität</b>								
<b>ItchRO(Obs) – Responder<sup>1)</sup></b>	<i>N<sup>2)</sup></i>	<i>Personen mit Ereignis, n (%)</i>		<i>N</i>	<i>Personen mit Ereignis, n (%)</i>		<i>RR [95%-KI]; p-Wert<sup>3)</sup></i>	
PFIC-Population	33	21 (63,6)		31	8 (25,8)		2,47 [1,29; 4,72]; 0,002	↑
Gesamtpopulation	47	k. A.		46	k. A.		k. A.	n. b.
<b>Körperliche Entwicklung: Körpergröße (z-Score)</b>	<i>n/N</i>	<i>Baseline MW (SD)</i>	<i>Änderung Woche 18–26 zu Baseline LS-MW (SE)</i>	<i>n/N</i>	<i>Baseline MW (SD)</i>	<i>Änderung Woche 18–26 zu Baseline LS-MW (SE)</i>	<i>LS-MWD [95%-KI]; p-Wert</i>	
PFIC-Population	32/33	-2,08 (1,31)	0,08 (0,09)	28/31	-2,06 (1,48)	-0,13 (0,09)	0,21 [-0,04; 0,45]; 0,09	↔
Gesamtpopulation	47	k. A.	k. A.	46	k. A.	k. A.	k. A.	n. b.
<b>Körperliche Entwicklung: Körpergewicht (z-Score)</b>	<i>n/N</i>	<i>Baseline MW (SD)</i>	<i>Änderung Woche 18–26 zu Baseline LS-MW (SE)</i>	<i>n/N</i>	<i>Baseline MW (SD)</i>	<i>Änderung Woche 18–26 zu Baseline LS-MW (SE)</i>	<i>LS-MWD [95%-KI]; p-Wert</i>	
PFIC-Population	32/33	-1,75 (1,31)	0,35 (0,07)	28/31	-1,28 (1,33)	0,12 (0,08)	0,23 [0,01; 0,44]; 0,04 <sup>4)</sup>	↑
Gesamtpopulation	47	k. A.	k. A.	46	k. A.	k. A.	k. A.	n. b.

Studie MARCH-PFIC Darstellung der Ergebnisse	Maralixibat		Placebo		Maralixibat vs. Placebo	Effekt
<b>Sicherheit<sup>5)</sup></b>						
<b>Schwere UE<sup>6)</sup></b>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%-KI]; p-Wert<sup>7)</sup></i>	
PFIC-Population	33	1 (3,0)	31	1 (3,2)	0,94 [0,06; 14,38]; 0,96	↔
Gesamtpopulation	47	3 (6,4)	46	3 (6,5)	k. A.	n. b. <sup>8)</sup>
<b>SUE<sup>6)</sup></b>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%-KI]; p-Wert<sup>7)</sup></i>	
PFIC-Population	33	3 (9,1)	31	3 (9,7)	0,94; [0,20; 4,31] 0,94	↔
Gesamtpopulation	47	5 (10,6)	46	3 (6,5)	k. A.	n. b. <sup>8)</sup>
<b>UE, das zum Abbruch der Studienmedikation führte<sup>6)</sup></b>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>N</i>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%-KI]; p-Wert<sup>7)</sup></i>	
PFIC-Population	33	0	31	0	0,94 [0,02; 46,04] n. b.	↔
Gesamtpopulation	47	1 (2,1)	46	0	k. A.	n. b. <sup>8)</sup>

<sup>1)</sup> Verbesserung  $\geq 1$  Punkt im Vergleich zu Baseline oder Schweregrad-Score  $\leq 1$  innerhalb der Wochen 15-26; es werden nur morgendliche Werte berücksichtigt.

<sup>2)</sup> Rücklaufquote unklar. Aus allgemeinen Angaben zum ItchRO-Tagebuch lässt sich eine ausreichende Datenvollständigkeit für den ItchRO(Obs) im gesamten Erhebungszeitraum ableiten.

<sup>3)</sup> P-Wert berechnet mittels Barnard's exaktem Test.

<sup>4)</sup> Hedges g nicht größer als die Relevanzschwelle 0,2.

<sup>5)</sup> Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind in Kapitel 3.5 dargestellt.

<sup>6)</sup> Es wurden studieneigene Kriterien zur Schweregradeinteilung verwendet („mild“, „moderat“, „schwer“).

<sup>7)</sup> CMH-Test, post hoc berechnet. Strata unklar.

<sup>8)</sup> Es wurden keine Effektschätzer (mit dazugehörigem p-Wert) anhand des Dossiers vorgelegt. Aufgrund der Anzahl an Ereignissen wird von keinem statistisch signifikanten Unterschied zwischen den Behandlungsarmen ausgegangen.

Abkürzungen: CMH: Cochran-Mantel-Haenszel; ItchRO(Obs): Itch Reported Outcome (Observer Version); k. A.: keine Angabe; KI: Konfidenzintervall; LS: Kleinste Quadrate; MW: Mittelwert; MWD: Mittelwertdifferenz; n. b.: nicht bewertbar; PFIC: Progressive familiäre intrahepatische Cholestase; RR: Relatives Risiko; SD: Standardabweichung; SE: Standardfehler; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Tabelle 25: Zusammenfassende Darstellung der Ergebnisse der Studie RISE

Studie RISE Darstellung der Ergebnisse	Maralixibat N = 4	
<b>Mortalität</b>	<i>N<sup>1)</sup></i>	<i>Personen mit Ereignis, n (%)</i>
Todesfälle	4	0
<b>Sicherheit</b>	<i>N<sup>1)</sup></i>	<i>Personen mit Ereignis, n (%)</i>
Schwere UE <sup>2)</sup>	4	0
SUE	4	1 (25,0)
UE, das zum Abbruch der Studienmedikation führte	4	0

<sup>1)</sup> Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden.

<sup>2)</sup> CTCAE-Version 5.0.

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

## Referenzen

1. **Alam S, Lal BB.** Recent updates on progressive familial intrahepatic cholestasis types 1, 2 and 3: Outcome and therapeutic strategies. *World J Hepatol* 2022;14(1):98-118.
2. **Baker A, Kerkar N, Todorova L, Kamath BM, Houwen RHJ.** Systematic review of progressive familial intrahepatic cholestasis. *Clin Res Hepatol Gastroenterol* 2019;43(1):20-36.
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: 11.09.2024]. URL: <https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>.
4. **Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E.** Progressive familial intrahepatic cholestasis. *Orphanet Journal of Rare Diseases* 2009;4(1):1.
5. **European Medicines Agency (EMA).** Livmarli (International non-proprietary name: maralixibat): European public assessment report EMEA/H/C/005857/II/0003/G; variation report [online]. 30.05.2024. Amsterdam (NED): EMA. [Zugriff: 15.07.2024]. URL: [https://www.ema.europa.eu/en/documents/variation-report/livmarli-h-c-005857-ii-0003-g-epar-assessment-report-variation\\_en.pdf](https://www.ema.europa.eu/en/documents/variation-report/livmarli-h-c-005857-ii-0003-g-epar-assessment-report-variation_en.pdf).
6. **European Medicines Agency (EMA).** Livmarli: Zusammenfassung der Merkmale des Arzneimittels (EPAR Product information) [online]. Amsterdam (NED): EMA; 2024. [Zugriff: 12.09.2024]. URL: [https://www.ema.europa.eu/de/documents/product-information/livmarli-epar-product-information\\_de.pdf](https://www.ema.europa.eu/de/documents/product-information/livmarli-epar-product-information_de.pdf).
7. **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.
8. **Le Pors C, Talagas M, Abasq-Thomas C, Henry S, Misery L, Roue JM.** What do we know about pruritus in very young infants? A literature review. *Cells* 2021;10(10).
9. **Mapi Research Trust.** Scaling and scoring for the acute and standard versions of the Pediatric Quality of Life Inventory PedsQL; version 21.3 [online]. In: ProQuolid. Stand: 03.2023. Lyon (FRA): Mapi Research Trust; 2023. [Zugriff: 18.10.2024]. URL: <https://www.pedsqgl.org/PedsQL-Scoring.pdf>.
10. **Miethke AG, Moukarzel A, Porta G, Covarrubias Esquer J, Czubkowski P, Ordonez F, et al.** Maralixibat in progressive familial intrahepatic cholestasis (MARCH-PFIC): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Gastroenterol Hepatol* 2024;9(7):620-631.
11. **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 (RISE); interim clinical study report [unveröffentlicht]. 2022.
12. **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 (RISE); statistical analysis plan [unveröffentlicht]. 2022.

13. **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 (RISE); study protocol and protocol amendments, version 5 [unveröffentlicht]. 2022.
14. **Mirum Pharmaceuticals.** Randomized double-blind placebo-controlled phase 3 study to evaluate the efficacy and safety of maralixibat in the treatment of subjects with progressive familial intrahepatic cholestasis (PFIC) – MARCH-PFIC; clinical study report [unveröffentlicht]. 2023.
15. **Mirum Pharmaceuticals.** Randomized double-blind placebo-controlled phase 3 study to evaluate the efficacy and safety of maralixibat in the treatment of subjects with progressive familial intrahepatic cholestasis (PFIC) – MARCH-PFIC; protocol [unveröffentlicht]. 2022.
16. **Mirum Pharmaceuticals.** Randomized double-blind placebo-controlled phase 3 study to evaluate the efficacy and safety of maralixibat in the treatment of subjects with progressive familial intrahepatic cholestasis (PFIC) – MARCH-PFIC; statistical analysis plan [unveröffentlicht]. 2022.
17. **Mirum Pharmaceuticals International.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 B: Maralixibat (Livmarli), zur Behandlung der progressiven familiären intrahepatischen Cholestase (PFIC) bei Patienten ab dem Alter von 3 Monaten; medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 26.07.2024.
18. **Verkade HJ, Thompson RJ, Arnell H, Fischler B, Gillberg PG, Mattsson JP, et al.** Systematic review and meta-analysis: partial external biliary diversion in progressive familial intrahepatic cholestasis. *J Pediatr Gastroenterol Nutr* 2020;71(2):176-183.
19. **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.
20. **World Health Organization (WHO).** Child growth standards [online]. Genf (SUI): WHO. [Zugriff: 11.09.2024]. URL: <https://www.who.int/tools/child-growth-standards>.

## Anhang

### sBA

Für die Studie MARCH-PFIC werden die Auswertungen der absoluten mittleren Veränderung der sBA von Woche 18, 22 und 26 gegenüber Baseline berücksichtigt. In der Kontrollgruppe zeigt sich ein leichter Anstieg der sBA-Werte gegenüber Baseline, während sich in der Interventionsgruppe eine ausgeprägte Reduktion zeigt. Der Unterschied zwischen den Behandlungsgruppen ist statistisch signifikant. Das Verzerrungspotential für den Endpunkt wird mit unklar bewertet, da auf Studienebene für die PFIC-Population von einem unklaren Verzerrungspotential ausgegangen wird.

*Tabelle 26: Veränderung der sBA zu Woche 18–26 gegenüber Baseline; Studie MARCH-PFIC, ITT-Population*

Studie MARCH-PFIC sBA ( $\mu\text{mol/l}$ )	PFIC-Population		Gesamtpopulation	
	MRX N = 33	PBO N = 31	MRX N = 47	PBO N = 46
Baseline n (%) MW (SD)	31 (93,9) 254,33 (139,52)	31 (100) 272,30 (147,40)	k. A.	k. A.
Veränderung zu Woche 18–26 <sup>1)</sup> n (%) LS-MW (SE)	29 (87,9) -157,49 (21,36)	28 (90,3) 2,91 (22,63)	k. A.	k. A.
LS-MWD [95%-KI] <sup>1)</sup> p-Wert SMD [95%-KI] <sup>2)</sup>	-160,40 [-220,84; -99,97] < 0,0001 -1,33 [-1,88; -0,78]		k. A.	

<sup>1)</sup> Schätzer aus einem MMRM mit „Veränderung zu Baseline“ als abhängiger Variable und den festen kategoriellen Effekten „Behandlungsgruppe“, „Zeitpunkt“, „PFIC-Typ“ und „Interaktion zwischen Behandlungsgruppe und Zeitpunkt“ und den kontinuierlichen festen Kovariaten „Mittlerer Baseline-Score“ und „Interaktion zwischen Baseline-Score und Zeitpunkt“.

<sup>2)</sup> SMD nach Hedges' g.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Kleinste Quadrate; MMRM: Mixed Model for Repeated Measures; MRX: Maralixibat; MW: Mittelwert; MWD: Mittelwertdifferenz; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; sBA: Serum-Gallensäure; SD: Standardabweichung, SE: Standardfehler; SMD: Standardisierte Mittelwertdifferenz.

In der Interventionsgruppe erfüllten 45,5 % der Personen das Responder-Kriterium, in der Kontrollgruppe waren es lediglich 6,4 %. Der Unterschied zwischen den Gruppen ist statistisch signifikant.

*Tabelle 27: Responderanalysen zur Verbesserung (Reduktion) der sBA; Studie MARCH-PFIC, ITT-Population*

Studie MARCH-PFIC sBA	PFIC-Population			Gesamtpopulation	
	MRX N = 33	PBO N = 31	RR [95%-KI]; p-Wert <sup>1)</sup>	MRX N = 47	PBO N = 46
sBA-Responder, n (%) <sup>2)</sup>	15 (45,5)	2 (6,4)	7,05 [1,75; 28,33]; 0,0004	k. A.	k. A.

<sup>1)</sup> P-Wert berechnet mittels Barnard's exaktem Test.

<sup>2)</sup> Responder waren definiert als Personen mit einem mittleren sBA-Level < 102  $\mu\text{mol/l}$  (nur, wenn sBA-Level zu Baseline  $\geq 102 \mu\text{mol/l}$  war) oder mind. 75 % mittlere relative Reduktion von Baseline. Eine Person galt als Non-Responder, wenn der Baseline-sBA-Wert fehlte oder die sBA-Werte zu allen Zeitpunkten (Woche 18, 22 und 26) fehlten.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MRX: Maralixibat; PBO: Placebo; PFIC: Progressive familiäre intrahepatische Cholestase; RR: Relatives Risiko; sBA: Serum-Gallensäure.

Für die Studie RISE erfolgte für die sBA-Werte eine rein deskriptive Auswertung. Zum Interim-Datenschnitt liegen ausschließlich Ergebnisse für die Kernstudie vor. Zu Woche 13 (Ende der Kernstudie) zeigt sich deskriptiv eine Reduktion gegenüber Baseline. Responderanalysen waren für sBA in der Studie RISE nicht vorgesehen.

*Tabelle 28: Veränderung der sBA zu Woche 13 gegenüber Baseline; Studie RISE, Sicherheitspopulation*

Studie RISE sBA (ergänzend dargestellt)	MRX N = 4	
	Beobachteter Wert	Veränderung gegenüber Baseline
<i>Baseline</i>		
n (%)	4 (100)	-
MW (SD)	211,72 (117,13)	
Median (min; max)	220,34 (60,9; 345,3)	
<i>Woche 13</i>		
n (%)	4 (100)	4 (100)
MW (SD)	139,89 (89,36)	-71,84 (179,86)
Median (min; max)	167,00 (13,2; 212,4)	-18,49 (-332,1; 81,7)

Abkürzungen: MRX: Maralixibat; MW: Mittelwert; sBA: Serum-Gallensäure; SD: Standardabweichung.