

# 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: Seladelpar**

Datum der Veröffentlichung: 16. Juni 2025

**Version 1.0**

## Inhaltsverzeichnis

Tabellenverzeichnis .....	3
Abbildungsverzeichnis .....	5
Abkürzungsverzeichnis .....	6
Hintergrund .....	8
1 Fragestellung .....	9
2 Auswahl und Methodik der relevanten Studien .....	10
2.1 Studienbasis für die Nutzenbewertung .....	10
2.2 Beschreibung der eingeschlossenen Studien .....	10
2.3 Endpunkte .....	16
2.3.1 Mortalität .....	17
2.3.2 Morbidität .....	17
2.3.3 Lebensqualität .....	27
2.3.4 Sicherheit .....	28
2.3.5 Erhebungszeitpunkte .....	30
2.4 Statistische Methoden .....	31
2.5 Verzerrungspotential auf Studien- und Endpunktebene .....	34
3 Ergebnisse der eingeschlossenen Studie .....	37
3.1 Studiencharakteristika und Studienmedikation .....	37
3.2 Mortalität .....	42
3.3 Morbidität .....	42
3.4 Lebensqualität .....	48
3.5 Sicherheit .....	50
4 Diskussion der Methodik und Ergebnisse .....	55
4.1 Zulassungsstatus und Zulassungspopulation von Seladelpar .....	55
4.2 Design und Methodik der Studie .....	56
4.3 Mortalität .....	58
4.4 Morbidität .....	58
4.5 Lebensqualität .....	60
4.6 Sicherheit .....	60
5 Anforderung an eine qualitätsgesicherte Anwendung .....	62
6 Zusammenfassung der Nutzenbewertung .....	63
Referenzen .....	67
Anhang .....	69

## Tabellenverzeichnis

Tabelle 1: Übersicht über die Studienbasis .....	10
Tabelle 2: Charakterisierung der Studie RESPONSE.....	10
Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie RESPONSE... 14	
Tabelle 4: Charakterisierung der Intervention der Studie RESPONSE .....	15
Tabelle 5: Zusammenfassung der Endpunktbewertung der Studie RESPONSE.....	16
Tabelle 6: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie RESPONSE .....	30
Tabelle 7: Verzerrungspotential der Studie RESPONSE .....	34
Tabelle 8: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie RESPONSE .....	34
Tabelle 9: Allgemeine Angaben; Studie RESPONSE (Ende der Studie: 11.08.2023) .....	37
Tabelle 10: Charakterisierung der Studienpopulation; Studie RESPONSE, ITT-Population.....	38
Tabelle 11: Veränderung der Pruritus NRS zu Monat 6 gegenüber Baseline auf Basis der gemittelten 7-Tage-Intervalle, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 .....	42
Tabelle 12: Veränderung der Pruritus NRS zu Monat 12 gegenüber Baseline auf Basis der gemittelten 7-Tage-Intervalle (ergänzend dargestellt), MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023.....	43
Tabelle 13: Pruritus-NRS-Ansprechen zu Monat 6 gegenüber Baseline bei Teilnehmenden mit Pruritus NRS > 0 zu Baseline (auf Basis der gemittelten 7-Tage-Intervalle), Responderanalysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 ..	44
Tabelle 14: Pruritus-NRS-Ansprechen zu Monat 12 gegenüber Baseline bei Teilnehmenden mit Pruritus NRS > 0 zu Baseline (auf Basis der gemittelten 7-Tage-Intervalle) (ergänzend dargestellt), Responderanalysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 .....	44
Tabelle 15: Veränderung der 5-D Itch zu Monat 12 gegenüber Baseline, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 .....	45
Tabelle 16: Veränderung des PBC-40 zu Monat 12 gegenüber Baseline, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 .....	48
Tabelle 17: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase, Studie RESPONSE, Sicherheitspopulation, Datenschnitt: 11.08.2023 .....	51
Tabelle 18: UE mit Inzidenz $\geq 10\%$ oder bei $\geq 10$ Patienten sowie mindestens $1\%$ in mindestens einem Studienarm nach Systemorganklasse und Preferred Term in der Studie RESPONSE, Datenschnitt: 11.08.2023 .....	51
Tabelle 19: UE von besonderem Interesse nach Systemorganklasse und Preferred Term in der Studie RESPONSE, Sicherheitspopulation, Datenschnitt: 11.08.2023 .....	52
Tabelle 20: Zusammenfassende Darstellung der Ergebnisse der Studie RESPONSE (ITT-Population).....	63

Tabelle 21: Biochemisches Ansprechen zu Monat 12, Responderanalysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 .....	69
Tabelle 22: PGI-S zu Monat 12; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023 ....	70
Tabelle 23: Veränderung des PGI-C zu Monat 12 gegenüber Monat 1, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023.....	70
Tabelle 24: Hospitalisierungen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023....	71

## **Abbildungsverzeichnis**

Abbildung 1: Schematische Darstellung des Studienablaufs der Studie RESPONSE [3] ..... 14

## Abkürzungsverzeichnis

AESI	Unerwünschtes Ereignis von besonderem Interesse (Adverse Event of Special Interest)
ALP	Alkalische Phosphatase
ALT	Alanin-Aminotransferase
AMA	Antimitochondriale Antikörper
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
APAC	Asien-Pazifik (Asia-Pacific)
AST	Aspartat-Aminotransferase
AUC	Fläche unter der Kurve (Area Under the Curve)
CDISC	Clinical Data Interchange Standards Consortium
CERC	Clinical Event Review Committee
CMH	Cochran-Mantel-Haenszel
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
EG	Europäische Gemeinschaft
eGFR	Geschätzte glomeruläre Filtrationsrate (Estimated Glomerular Filtration Rate)
ELF	Fortgeschrittene Leberfibrose (Enhanced Liver Fibrosis)
EMA	European Medicines Agency
EMEA	Europa, Naher Osten und Afrika (Europe, the Middle East, and Africa)
EOT	Ende der Behandlung (End of Treatment)
EPAR	European Public Assessment Report
ET	Early Termination
FAS	Full-Analysis-Set
FDA	U. S. Food and Drug Administration
FMQ	FDA MedDRA Queries
FU	Follow-Up
G-BA	Gemeinsamer Bundesausschuss
GGT	Gamma-Glutamyltransferase
GKV	Gesetzliche Krankenversicherung
INR	International Normalized Ratio
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to Treat
k. A.	keine Angabe
KI	Konfidenzintervall
kPa	Kilopascal
LLN	Unterer Normwert (Lower Limit of Normal)
LS	Least Squares
M	Monat
MedDRA	Medical Dictionary for Regulatory Activities

MELD	Model for End-Stage Liver Disease
MMRM	Gemischtes Modell mit Messwiederholungen (Mixed Effect Model for Repeated Measurements)
MSPN	Moderater-bis-Schwerer-Pruritus-Numerical-Rating-Scale (Moderate-to-Severe-Pruritus-Numerical-Rating-Scale)
MW	Mittelwert
N	Anzahl
n. a.	nicht anwendbar
n. b.	nicht bewertbar
NRS	Numerical Rating Scale
OCA	Obeticholsäure (Obeticholic Acid)
OR	Odds Ratio
PBC	Primär biliäre Cholangitis
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PT	Preferred Term
pU	pharmazeutischer Unternehmer
QoL	Quality of Life
R	Randomisierung
RD	Risikodifferenz
RR	Relatives Risiko
SAP	Statistischer Analyseplan
SAS	Safety-Analysis-Set
SD	Standardabweichung (Standard Deviation)
SE	Standardfehler (Standard Error)
SGB	Sozialgesetzbuch
SMD	Standardisierte Mittelwertdifferenz
SMQ	Standardised MedDRA Queries
SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis
T	Tag
TEAE	Treatment Emergent Adverse Event
UDCA	Ursodeoxycholsäure (Ursodeoxycholic Acid)
UE	Unerwünschtes Ereignis
UK	United Kingdom
ULN	Oberer Normwert (Upper Limit of Normal)
UNS-Visite	Unscheduled Visit
VerfO	Verfahrensordnung des G-BA

## Hintergrund

Seladelpar 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 Seladelpar 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 Seladelpar in seiner Sitzung am 11. Juni 2025 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 17. März 2025 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 16. Juni 2025 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

Seladelpar (Lyvdelzi<sup>®</sup>, ehemals Seladelpar Gilead) 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 [10]:

Behandlung der primär biliären Cholangitis (PBC) in Kombination mit Ursodeoxycholsäure (UDCA) bei Erwachsenen, die nicht ausreichend auf UDCA allein ansprechen, oder als Monotherapie bei Patientinnen und Patienten, die UDCA nicht vertragen.

Gemäß Fachinformation beträgt die empfohlene Dosierung 10 mg einmal täglich.

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studienname (Nummer)	Studie mit Dossier vom pU eingereicht	Studie vom pU als relevant für die Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzenbewertung	Ausschlussgrund
<b>Studien zum Wirkstoff</b>				
RESPONSE (CB8025-32048) <sup>1)</sup>	Ja	Ja	Ja	-

<sup>1)</sup> Pivotalstudie gemäß EPAR. Teilnehmende der Studie RESPONSE und aus anderen Studien zu Seladelpar hatten die Möglichkeit im Anschluss an der offenen, einarmigen Langzeitnachbeobachtungsstudie (LTE) ASSURE (CB8025-31731-RE) teilzunehmen. Aufgrund des einarmigen Studiendesigns der LTE bei Vorliegen eines geeigneten RCT mit einjähriger Beobachtungsdauer wird die Studie ASSURE nicht zur Nutzenbewertung herangezogen.

Abkürzungen: EPAR: European Public Assessment Report; LTE: Langzeitnachbeobachtungsstudie; pU: pharmazeutischer Unternehmer; RCT: Randomized Controlled Trial.

Die für die Nutzenbewertung herangezogene Studie entspricht derjenigen Studie, auf der die Einschätzungen des pU zum Zusatznutzen basieren.

#### Zur Nutzenbewertung für Seladelpar herangezogene Unterlagen

- Dossier des pU zu Seladelpar [8,9]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [7]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der pivotalen Studie RESPONSE (CB8025-32048) [3,4,5,6]
- Fachinformation zu Seladelpar [10]

### 2.2 Beschreibung der eingeschlossenen Studien

Die Nutzenbewertung von Seladelpar im vorliegenden Anwendungsgebiet basiert auf der Zulassungsstudie RESPONSE (CB8025-32048). Die Studie und die Intervention werden in Tabelle 2 bis Tabelle 4 charakterisiert.

Tabelle 2: Charakterisierung der Studie RESPONSE

Charakteristikum	Beschreibung
<b>Design und Studienablauf</b>	<p>RESPONSE ist eine multizentrische, doppelblinde, randomisierte, Placebo-kontrollierte, Phase-III-Parallelgruppenstudie zur Untersuchung der Wirksamkeit, Verträglichkeit und Pharmakokinetik von Seladelpar bei Erwachsenen mit PBC.</p> <p>Die Randomisierung erfolgte stratifiziert im Verhältnis 2:1 (Seladelpar:Placebo) anhand der folgenden Stratifizierungsfaktoren:</p> <ul style="list-style-type: none"> <li>• ALP-Level (ALP &lt; 350 U/L vs. ALP ≥ 350 U/L)</li> <li>• Pruritus (NRS &lt; 4 vs. NRS ≥ 4)</li> </ul>

Charakteristikum	Beschreibung
	<p><u>Die Studie ist in folgende Abschnitte unterteilt (siehe auch Abbildung 1):</u></p> <ul style="list-style-type: none"> <li>• Screening: bis zu 3 Wochen<sup>1)</sup></li> <li>• Run in: 2 Wochen</li> <li>• Behandlungszeitraum: 12 Monate</li> <li>• Sicherheitsnachbeobachtung: 2 Wochen (+ 3 Tage)</li> </ul> <p>Die Gesamtdauer beträgt somit ca. 14 Monate. Nach Abschluss des 12-monatigen Behandlungszeitraums hatten die Teilnehmenden die Möglichkeit, an einer offenen Langzeitnachbeobachtungsstudie (ASSURE) teilzunehmen. Für Patientinnen und Patienten, die in die Studie ASSURE übergangen, entfiel die reguläre Sicherheitsnachbeobachtung der Studie RESPONSE.</p> <p>Die Intervention wurde bis zum Abbruch der Therapie, dem Austritt aus der Studie, dem Eintreten eines klinischen PBC-Ereignisses<sup>2)</sup> oder bis Monat 12/EOT durchgeführt, je nachdem, was früher eintrat.</p> <p>Teilnehmende, bei denen ein klinisches PBC-Ereignis<sup>2)</sup> auftrat, wurden aus der Studie ausgeschlossen und absolvierten eine vorzeitige Abschlussvisite.</p> <p>Teilnehmende, die die Studienteilnahme aus einem anderen Grund als einem PBC-Ereignis vorzeitig abbrechen wollten, wurden gebeten, die Studienteilnahme ohne Studienmedikation fortzuführen. Wurde dies abgelehnt und die Studienteilnahme zu einem beliebigen Zeitpunkt nach Tag 1 beendet, wurde eine Early-Termination(ET)-Visite durchgeführt</p> <p>Die Studie ist abgeschlossen.</p>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• 18 bis einschließlich 75 Jahre alt</li> <li>• Diagnose einer PBC, die auf zwei der folgenden Kriterien beruht: <ul style="list-style-type: none"> <li>◦ ALP-Wert über 1,0 x ULN seit mindestens 6 Monaten;</li> <li>◦ Positiver AMA-Titer (&gt; 1:40 in der Immunfluoreszenz oder M2 positiv im Enzymimmunoassay oder positive PBC-spezifische antinukleäre Antikörper);</li> <li>◦ Dokumentierte Leberbiopsieergebnisse, die auf PBC hindeuten.</li> </ul> </li> <li>• UDCA-Behandlung während der letzten 12 Monate (stabile Dosis seit &gt; 3 Monaten vor dem Screening) ODER Unverträglichkeit von UDCA (letzte UDCA-Dosis &gt; 3 Monate vor dem Screening).</li> <li>• Vom Zentrallabor beim Screening gemessene Laborparameter<sup>3)</sup> <ul style="list-style-type: none"> <li>◦ ALP <math>\geq 1,67 \times \text{ULN}</math>,</li> <li>◦ AST <math>\leq x \text{ ULN}</math>,</li> <li>◦ ALT <math>\leq 3 \times \text{ULN}</math>,</li> <li>◦ Gesamt-Bilirubin <math>\leq 2 \times \text{ULN}</math>,</li> <li>◦ eGFR <math>&gt; 45 \text{ mL/min/1,73 m}^2</math>,<sup>4)</sup></li> <li>◦ INR <math>&lt; 1,1 \times \text{ULN}</math>,<sup>5)</sup></li> <li>◦ Thrombozytenzahl <math>\geq 100 \times 10^3/\mu\text{L}</math>.</li> </ul> </li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Fortgeschrittene PBC im Sinne der Rotterdam-Kriterien (Albumin <math>&lt; \text{LLN}</math> UND Gesamt-Bilirubin <math>&gt; 1,0 \times \text{ULN}</math>)</li> <li>• Akute und/oder historische weitere Erkrankungen <ul style="list-style-type: none"> <li>◦ Vorliegen einer klinisch bedeutsamen dekompensierten Lebererkrankung, einschließlich <ul style="list-style-type: none"> <li>▪ Lebertransplantation in der Vorgeschichte, aktuelle Aufnahme in die Lebertransplantationsliste oder aktueller MELD-Wert <math>\geq 12</math>;</li> </ul> </li> </ul> </li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>▪ Komplikationen der portalen Hypertension, einschließlich bekannter Ösophagusvarizen, früherer Varizenblutungen oder damit zusammenhängende Eingriffe, Aszites und hepatischer Enzephalopathie;</li> <li>▪ Zirrhose mit Komplikationen, einschließlich Vorgeschichte oder Vorliegen einer spontanen bakteriellen Peritonitis, eines hepatozellulären Karzinoms oder eines hepatorenenalen Syndroms.</li> <li>○ Vorliegen anderer chronischer Lebererkrankungen: <ul style="list-style-type: none"> <li>▪ Aktuelle Merkmale einer Autoimmunhepatitis</li> <li>▪ Primäre sklerosierende Cholangitis</li> <li>▪ Vorgeschichte oder klinische Anzeichen einer alkoholischen Lebererkrankung, eines Alpha-1-Antitrypsin-Mangel, einer nicht-alkoholischen Steatohepatitis, eines Gilbert-Syndroms mit erhöhtem Gesamt-Bilirubin oder einer Hämochromatose</li> <li>▪ Hepatitis B oder Hepatitis C</li> <li>▪ Vorgeschichte, Nachweis oder starker Verdacht auf hepatobiliäre Malignität</li> </ul> </li> <li>○ Klinisch bedeutsamer Alkoholkonsum<sup>6)</sup>, oder die Unfähigkeit, den Alkoholkonsum zuverlässig zu quantifizieren</li> <li>• Die Behandlung mit verschiedenen Medikamenten zu unterschiedlichen Zeitpunkten vor Studieneintritt war ausgeschlossen: <ul style="list-style-type: none"> <li>○ während der letzten 4 Wochen vor dem Screening: pruritusstillende Medikamente, außer die Dosis war über diesen Zeitraum stabil</li> <li>○ während der letzten 6 Wochen vor dem Screening: OCA und Fibrate<sup>7)</sup></li> <li>○ während der letzten 8 Wochen vor dem Screening: Colchizin, Methotrexat, Azathioprin oder langfristige (&gt; 2 Wochen) systematische Kortikosteroide</li> <li>○ generell: <ul style="list-style-type: none"> <li>▪ Immunsuppressive Therapien<sup>8)</sup> (z. B. Cyclosporin, Tacrolimus, Anti-Tumornekrosefaktor oder andere immunsuppressive Biologika)</li> <li>▪ Andere Medikamente, die sich auf die Leber- oder Magen-Darm-Funktionen auswirken, wie z. B. die Absorption von Medikamenten oder Roux-en-Y-Magenbypass-Verfahren<sup>9)</sup>, können verboten sein und sollten von Fall zu Fall mit dem klinischen Monitor besprochen werden.<sup>8)</sup></li> </ul> </li> </ul> </li> <li>• Für Studienzentren in Deutschland: in Frage kommen für eine Behandlung mit OCA (bspw. Personen, bei denen keine unzureichende Wirksamkeit, keine medizinische Kontraindikation oder Unverträglichkeit gegenüber OCA bekannt ist oder die eine Behandlung mit OCA nicht ablehnen).</li> </ul>
<b>Intervention und Zahl der Patientinnen und Patienten</b>	Gescreent gesamt: N = 360 Randomisiert: N = 193 Seladelpar: N = 128 Placebo: N = 65
<b>Ort und Zeitraum der Durchführung; Datenschnitte</b>	<b>Studienzentren</b> 90 Studienzentren in 24 Ländern (Argentinien, Australien, Österreich, Belgien, Kanada, Chile, Tschechische Republik, Frankreich, Deutschland, Griechenland, Ungarn, Israel, Italien, Republik Korea, Mexiko, Neuseeland, Polen, Rumänien, Russland, Spanien, Schweiz, Türkei, Vereinigtes Königreich und Vereinigte Staaten von Amerika).

Charakteristikum	Beschreibung
	<b>Studienzeitraum</b> <ul style="list-style-type: none"> <li>• Erste Randomisierung: 21. April 2021</li> <li>• Letzter Patient, letzte Visite: 11. August 2023</li> </ul> <b>Datenschnitt</b> 11.08.2023 <sup>10)</sup>
<b>Endpunkte gemäß Studienprotokoll</b>	<b>Primärer Endpunkt</b> <ul style="list-style-type: none"> <li>• Kombiniertes biochemisches Ansprechen (definiert als ALP &lt; 1,67 × ULN, ALP-Reduktion ≥ 15 % und Gesamtbilirubin ≤ 1.0 × ULN) zu Monat 12</li> <li>• Unerwünschte Ereignisse</li> </ul> <b>Sekundäre Endpunkte</b> <ul style="list-style-type: none"> <li>• Kombiniertes biochemisches Ansprechen zu Monat 6 <ul style="list-style-type: none"> <li>◦ Anteil Personen mit ALP ≤ 1,0× ULN, &lt;1,67× ULN and ALP &lt; 1,5× ULN</li> </ul> </li> <li>• absolute und relative Veränderungen von ALP, ALT, AST, GGT, Bilirubin und 5'-Nukleotidase</li> <li>• UK-PBC- und GLOBE-Risiko-Scores</li> <li>• Pruritus: NRS</li> <li>• Klinische PBC-Ereignisse <ul style="list-style-type: none"> <li>◦ Todesfälle</li> <li>◦ Lebertransplantation</li> <li>◦ MELD-Score ≥ 15 bei mindestens 2 aufeinanderfolgenden Visiten</li> <li>◦ behandlungsbedürftiger Aszites</li> <li>◦ Hospitalisierung aufgrund definierter Ereignisse</li> </ul> </li> <li>• PBC-40 QoL</li> </ul> <b>Explorative Endpunkte</b> <ul style="list-style-type: none"> <li>• Pruritus: 5-D Itch, PBC-40 Juckreiz-Domäne, zusätzliche Pruritus-NRS-Analysen</li> <li>• Patienteneindruck des Pruritus: PGI-S, PGI-C</li> <li>• Fibrose (FibroScan®, ELF-Score)</li> <li>• Leberhistologie, weitere Biomarker und Laborparameter (u.a. C4, FGF19, CK18)</li> <li>• PBC-Ansprechkriterien (Barcelona, Paris I und II, Toronto I und II, Rotterdam)</li> </ul>

<sup>1)</sup> Angabe zur dreiwöchigen Screeningphase gemäß Protokoll und Studienbericht. In Modul 4 ist für das Screening ein abweichender Zeitraum von bis zu 3 Monaten angegeben. Daher beträgt laut Modul 4 die Gesamtdauer der Studie ebenfalls abweichend ca. 16 statt ca. 14 Monate.

<sup>2)</sup> Als klinische PBC-Ereignisse waren definiert: Todesfälle, Lebertransplantation, MELD-Score ≥ 15 bei mindestens 2 aufeinanderfolgenden Visiten, behandlungsbedürftiger Aszites, Hospitalisierung aufgrund definierter Ereignisse. Die nähere Definition der klinischen PBC-Ereignisse ist Kapitel 2.3 zu entnehmen.

<sup>3)</sup> Prothrombinzeit, INR und Thrombozyten können bei der Screening-Visite vor Ort (also lokal) bestimmt werden, wenn die Prüffärztin/der Prüffarzt dies nach Rücksprache mit dem klinischen Monitor für notwendig hält, falls die zentral gemessenen Proben als ungültig erachtet werden.

<sup>4)</sup> Das Einschlusskriterium wurde mit Protokollversion 4 (09.02.2022, d. h. etwa ein Jahr nach der ersten Randomisierung) in Absprache mit der FDA von > 60 mL/min/1,73 m<sup>2</sup> auf ≥ 45 mL/min/1,73 m<sup>2</sup> gesenkt. Berechnet wurde die eGFR nach der Gleichung der Modification of Diet in Renal Disease Study.

<sup>5)</sup> Bei Patientinnen und Patienten, die eine Antikoagulationstherapie erhalten, muss der INR-Wert in dem Bereich gehalten werden, der für die Prophylaxe der spezifischen Erkrankung erforderlich ist.

<sup>6)</sup> Definiert als mehr als 2 Getränkeinheiten pro Tag (entspricht 20 g) bei Frauen und 3 Getränkeinheiten pro Tag (entspricht 30 g) bei Männern.

<sup>7)</sup> Das Ausschlusskriterium wurde mit Protokollversion 4 (09.02.2022) angepasst: Eine Behandlung mit Obeticholsäure oder Fibraten führte nun zum Ausschluss, wenn sie innerhalb von 6 Wochen (statt zuvor innerhalb von 3 Monaten) vor dem Screening erfolgte.

<sup>8)</sup> Ausschlusskriterium mit Protokollversion 3 (30.06.2021) eingeführt.

<sup>9)</sup> Das Ausschlusskriterium „Roux-en-Y-Magenbypass-Verfahren“ wurde mit Protokollversion 4 (09.02.2022) eingeführt.

<sup>10)</sup> Angabe gemäß Modul 4. In den Studienunterlagen findet sich keine explizite Nennung eines Datenschnitts. Laut EPAR erfolgte die Datenbanksperre am 30. August 2023.

Abkürzungen: ALP: Alkalische Phosphatase; ALT: Alanin-Aminotransferase; AMA: Antimitochondriale Antikörper; AST: Aspartat-Aminotransferase; eGFR: Geschätzte glomeruläre Filtrationsrate; ELF: Fortgeschrittene Leberfibrose; EOT: Ende der Behandlung; FDA: U. S. Food and Drug Administration; GGT: Gamma-Glutamyltransferase; INR: International Normalized Ratio; LLN: Unterer Normwert; MELD: Model for End-Stage Liver Disease; NRS: Numerical Rating Scale; PBC: Primär biliäre Cholangitis; PGI-C: Patient Global Impression of Change; PGI-S: Patient Global Impression of Severity; QoL: Quality of Life; UDCA: Ursodeoxycholsäure; UK: United Kingdom; ULN: Oberer Normwert.

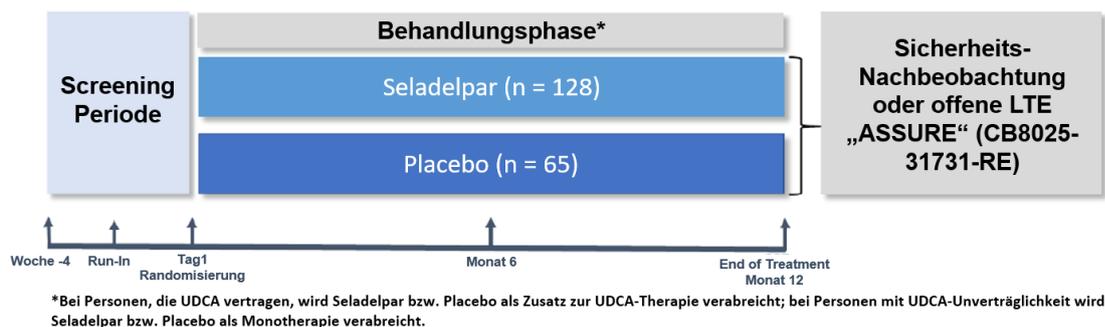


Abbildung 1: Schematische Darstellung des Studienablaufs der Studie RESPONSE [3]

### Protokolländerungen

Es wurden 3 globale Änderungen des Originalprotokolls vom 10.09.2020 vorgenommen. 2 dieser globalen Amendments lagen nach der Randomisierung der ersten Studienperson (21.04.2021). Zudem existieren für die Protokollversionen 2, 3 und 4 länderspezifische Amendments für Deutschland (Protokollversionen 2.1, 3.1 und 4.1). Diese enthalten eine zusätzliche Anpassung: Personen, die außerhalb der Studienteilnahme für eine Behandlung mit Obeticholsäure (OCA) infrage gekommen wären, wurden auf Empfehlung der Ethikkommission ausgeschlossen. Abgesehen davon sind die deutschen Protokollversionen gemäß pU inhaltlich identisch mit den globalen Studienprotokollen. Die globalen Originale liegen nicht vor. Von den beiden globalen Amendments nach der ersten Randomisierung enthält nur Amendment 4 für die Nutzenbewertung relevante Änderungen. Diese sind in Tabelle 3 dargestellt.

Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie RESPONSE

Amendment	Wesentliche Änderungen
Version 4 vom 09.02.2022, (Anzahl der bis dahin eingeschlossenen Personen: k. A.)	<ul style="list-style-type: none"> <li>• Nach einer Diskussion mit der FDA wurden die Einschlusskriterien für die eGFR von &gt; 60 mL/min/1,73 m<sup>2</sup> auf ≥ 45 mL/min/1,73 m<sup>2</sup> angepasst.</li> <li>• Kürzung des Zeitraums für die Sicherheitsnachbeobachtung von 1 Monat nach der letzten Studiendosis auf 14 Tage (+3).</li> <li>• Screening-Zeitfenster wurde verlängert, um mit der durchschnittlichen Zeit der Standorte für die Terminierung von Screening-Untersuchungen übereinzustimmen (von bis zu 2 auf bis zu 3 Wochen).</li> </ul>

Abkürzungen: eGFR: Geschätzte glomeruläre Filtrationsrate; FDA: U. S. Food and Drug Administration; k. A.: keine Angabe.

## Charakterisierung der Intervention

Tabelle 4: Charakterisierung der Intervention der Studie RESPONSE

Intervention	Kontrolle
<p><b>Seladelpar</b>  <u>Dosierung und Art der Anwendung</u></p> <ul style="list-style-type: none"> <li>• Anfangsdosis: einmal täglich peroral 10 mg Hartkapsel</li> <li>• Anwendungsdauer: 12 Monate</li> </ul>	<p><b>Placebo</b>  <u>Dosierung und Art der Anwendung</u>  <i>Entsprechend der Intervention, ohne den aktiven Wirkstoff.</i></p>
<p><u>Dosisanpassungen:</u>  Bei Personen, die spezifische Sicherheitsüberwachungskriterien<sup>1)</sup> erfüllen oder Verträglichkeitsprobleme zeigen, konnte in beiden Studienarmen eine Dosisreduktion auf 5 mg unter Beibehaltung der Verblindung vorgenommen werden. Präspezifizierte Vorgaben darüber, ob und inwiefern auf eine Dosisreduktion eine erneute Dosiserhöhung stattfinden konnte und/oder sollte, konnten dem Studienprotokoll nicht entnommen werden.  Teilnehmende, bei denen ein klinisch bedeutsames UE<sup>2)</sup> auftrat, das nach dem Urteil der Prüffärztin oder des Prüfarztes eine Dosisreduktion rechtfertigte, kamen ebenfalls für eine gemäß Studienunterlagen „ähnliche“<sup>3)</sup> verblindete Dosisreduktion in Frage.</p> <p><u>Unterbrechungen der Studienmedikation</u></p> <ul style="list-style-type: none"> <li>• Personen, bei denen die Studienmedikation aufgrund eines UE vorübergehend unterbrochen wurde, konnten die Anfangsdosis wieder aufnehmen, wenn das UE abgeklungen war und es gemäß Prüfpersonaleinschätzung in keinem Zusammenhang mit dem Studienmedikament stand.</li> <li>• Unterbrechungen konnten zudem erfolgen für erforderliche medizinische Maßnahmen, Krankheiten, Reisebeschränkungen oder andere Aktivitäten, welche die Teilnehmenden daran hätten hindern können, die Dosis gemäß Protokoll einzunehmen.</li> </ul> <p><u>UDCA</u>  Während der Studie wird die Studienmedikation bei Personen, die UDCA vertragen, als Zusatz zur UDCA-Therapie verabreicht; bei Personen mit UDCA-Unverträglichkeit wird Seladelpar bzw. Placebo als Monotherapie verabreicht.</p> <ul style="list-style-type: none"> <li>• <u>Dosierung:</u> oral (Hartkapseln) einmal oder mehrmals täglich in Dosishöhe wie vor Studienteilnahme und basierend auf dem klinischen Ermessen der Prüffärztin / des Prüfarztes.</li> </ul>	
<p><b>Nicht erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Obeticholsäure (OCA)</li> <li>• Fibrate (z. B. Bezafibrat, Fenofibrat, Elafibranor, Lanifibranor, Pemafibrat, Saroglitazar)</li> <li>• Colchicin, Methotrexat oder Azathioprin</li> <li>• Langfristige systemische Steroide (z. B. Prednison, Prednisolon, Budesonid) für &gt; 2 Wochen</li> <li>• Experimentelle oder nicht zugelassene Behandlungen für PBC oder verwandte Autoimmunkrankheiten</li> <li>• Immunsuppressive Therapien (z. B. Cyclosporin, Tacrolimus, Anti-TNF oder andere immunsuppressive Biologika)</li> <li>• Andere Medikamente, die die Leber- oder Magen-Darm-Funktionen beeinflussen, z. B. die Absorption von Medikamenten, können verboten sein und sollten von Fall zu Fall mit dem klinischen Monitor besprochen werden.</li> </ul> <p><b>Erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Erforderliche medikamentöse und therapeutische Interventionen zur Behandlung neuer oder bestehender medizinischer Erkrankungen.</li> <li>• Die symptomatische Behandlung der PBC (z. B. juckreizstillende Medikamente) sollte mit dem klinischen Monitor besprochen werden.</li> </ul>	

- <sup>1)</sup> Im Studienprotokoll werden Kriterien für die Sicherheitsüberwachung in Bezug auf Leber, Muskelverletzungen, Nieren und Bauchspeicheldrüse festgelegt. Dabei werden Grenzwerte für Laborparameter definiert sowie Empfehlungen gegeben, ab welchen Werten, ggf. unter Erfüllung zusätzlicher Kriterien, die Studienmedikation vorübergehend oder dauerhaft unterbrochen werden sollte.
- <sup>2)</sup> In den Studienunterlagen ist eine klinisch bedeutsame Anomalie definiert als eine bestätigte Anomalie, die gegenüber Baseline ausreichend verändert ist, so dass nach dem klinischen Urteil des Prüfers eine Änderung der Behandlung gerechtfertigt ist. Es ist nicht eindeutig ersichtlich, ob sich „Anomalien“ dabei auf UE im Allgemeinen beziehen oder nur auf anormale Laborwerte.
- <sup>3)</sup> Den Studienunterlagen konnte nicht entnommen werden, ob und inwiefern sich die Dosisreduktion aufgrund klinisch bedeutsamer UE möglicherweise von der Dosisreduktion aufgrund der Sicherheitskriterien unterschied.

Abkürzungen: OCA: Obeticholsäure; PBC: Primär biliäre Cholangitis; UDCA: Ursodeoxycholsäure; UE: Unerwünschtes Ereignis.

## 2.3 Endpunkte

In diesem Kapitel werden die Eignung der Endpunkte hinsichtlich Operationalisierung, Patientenrelevanz und Validität beurteilt. Dazu wurden das Herstellerdossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt. Alle Endpunkte der eingeschlossenen Studien (siehe Tabelle 2) wurden einer Prüfung unterzogen. Endpunkte, die in der nachfolgenden Tabelle 5 nicht gelistet sind, wurden weder vom pU noch im Rahmen der Nutzenbewertung als patientenrelevant bzw. bewertungsrelevant eingestuft. Tabelle 5 stellt das Ergebnis dieser Bewertung zusammenfassend dar.

*Tabelle 5: Zusammenfassung der Endpunktbewertung der Studie RESPONSE*

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Todesfälle	Mortalität	Ja	Ja
Biochemisches Ansprechen <sup>1)</sup>	Morbidity	Ja	Nein <sup>2)</sup>
Pruritus			
Pruritus NRS		Ja	Ja
5-D Itch		Ja	Ja
PGI-S		Ja	Nein <sup>3)</sup>
PGI-C		Ja	Nein <sup>4)</sup>
Fibrose		Ja	Nein
Klinische PBC-Ereignisse <sup>5)</sup>			
Lebertransplantation		Nein	Nein <sup>6)</sup>
MELD-Score $\geq 15$ <sup>7)</sup>		Nein	Ja
Behandlungsbedürftiger Aszites	Nein	Nein <sup>6)</sup>	
Hospitalisierung	Ja	Nein <sup>6)</sup>	
PBC-40	Lebensqualität	Ja	Ja
Unerwünschte Ereignisse <sup>1)</sup>	Sicherheit	Ja	Ja

<sup>1)</sup> Primärer Endpunkt

<sup>2)</sup> Der primäre Endpunkt wird aus Transparenzgründen im Anhang dargestellt.

<sup>3)</sup> Der PGI-S wird aus Transparenzgründen im Anhang dargestellt.

<sup>4)</sup> Der PGI-C wird aus Transparenzgründen im Anhang dargestellt.

- <sup>5)</sup> Todesfälle waren auch als ein klinisches PBC-Ereignis definiert. Im Rahmen der Nutzenbewertung erfolgt jedoch keine erneute Nennung dieses eigenständigen Mortalitätsendpunktes.
- <sup>6)</sup> Bei dem Endpunkt kann es sich in der vorliegenden Indikation grundsätzlich um einen patientenrelevanten Endpunkt handeln. Jedoch wird aufgrund wesentlicher Unklarheiten in Bezug zur Operationalisierung die Validität des Endpunktes als unklar eingestuft (siehe auch Kapitel 2.3.2). Aus Transparenzgründen erfolgt eine Darstellung im Anhang.
- <sup>7)</sup> Bei mindestens 2 aufeinanderfolgenden Visiten.

Abkürzungen: MELD: Model for End-Stage Liver Disease; NRS: Numerical Rating Scale; PBC: Primär biliäre Cholangitis; PGI-C: Patient Global Impression of Change; PGI-S: Patient Global Impression of Severity; pU: pharmazeutischer Unternehmer.

### **2.3.1 Mortalität**

#### **Todesfälle**

Der Endpunkt „Todesfälle“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Todesfälle jeglicher Ursache wurden über die Sicherheit kontinuierlich ab dem Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis zum Ende der Studienteilnahme, also bis zur letzten Studienvisite, erhoben.

##### *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

Die Erhebung des Endpunktes „Todesfälle“ wird als valide eingeschätzt.

### **2.3.2 Morbidität**

#### **Biochemisches Ansprechen**

Der kombinierte Endpunkt „Biochemisches Ansprechen“ wird aufgrund von fehlender unmittelbarer Patientenrelevanz nicht zur Nutzenbewertung herangezogen. Als primärer Endpunkt wird dieser jedoch aus Transparenzgründen im Anhang dargestellt.

#### Operationalisierung

##### *Beschreibung*

Das kombinierte biochemische Ansprechen auf die Behandlung zu Monat 12 gegenüber Baseline war als primärer Endpunkt definiert. Es handelt sich um einen kombinierten Endpunkt, der aus den folgenden Teilkomponenten besteht:

- ALP (Alkalische Phosphatase) < 1,67 × ULN (Upper Limit of Normal) und
- ALP-Reduktion ≥ 15 % und
- Gesamtbilirubin ≤ 1,0 × ULN.

Die Entnahme der Blutproben für die Laboruntersuchungen erfolgte bei den Studienvisiten jeweils vor dem Erhalt der Studiendosis. Teilnehmende sollten zum Zeitpunkt der Entnahme seit mindestens 8 Stunden nüchtern sein. Vergaß die Person nüchtern zu sein, wurde dies von der Prüfstelle vermerkt, die Laboruntersuchungen wurden jedoch regulär fortgesetzt.

Neben der Auswertung des kombinierten Endpunkts waren auch verschiedene Auswertungen der Einzelkomponenten vorgesehen. Die Auswertung des primären Endpunktes erfolgte über den Anteil der Responder gemäß den Kriterien für das kombinierte biochemische Ansprechen.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Der Endpunkt erfasst ausschließlich Befunde auf Basis von Laborparametern und ohne direkten Symptombefund, die als nicht unmittelbar patientenrelevant beurteilt werden. Geeignete Studien zur Surrogatvalidierung des biochemischen Ansprechens für patientenrelevante Endpunkte, wie „Tod“ oder „Lebertransplantation“ wurden nicht eingereicht.

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

#### Validität

Die vom pU eingereichten Studien zur Untersuchung eines Zusammenhangs zwischen den Parametern Gesamtbilirubin und Alkalische Phosphatase mit dem Gesamtüberleben bzw. dem transplantationsfreien Überleben werden für eine Surrogatvalidierung als nicht geeignet angesehen. Gründe sind u. a. das Vorliegen von Studien mit einarmigem Studiendesign bzw. prospektive Kohortenstudien und eine nicht gegebene Methodenspezifität. Todesfälle sowie das Ereignis „Lebertransplantation“ werden in der Studie RESPONSE außerdem bereits als eigenständige Endpunkte erhoben.

Eine konkrete Begründung, weshalb die Teilkomponente „Alkalische Phosphatase“ in zwei unterschiedlichen Operationalisierungen („ALP < 1,67 × ULN“ und „ALP-Reduktion ≥ 15 %“) Bestandteil des kombinierten Endpunktes ist, konnte nicht identifiziert werden.

Aufgrund der unklaren Patientenrelevanz und der fehlenden Surrogatvalidierung wird der primäre Endpunkt nicht in der Nutzenbewertung berücksichtigt, aber aus Transparenzgründen im Anhang abgebildet.

#### **Pruritus**

Die folgenden 4 Endpunkte erfassen das Symptom „Juckreiz“ anhand von unterschiedlichen, aber teilweise sehr ähnlichen, Operationalisierungen. Dadurch können sich Erhebungsinhalte teilweise bis erheblich überschneiden. In der vorliegenden Nutzenbewertung werden die verschiedenen Operationalisierungen daher mit der Endpunktbezeichnung „Pruritus“ zusammengefasst und mögliche Mehrfacherhebungen werden bei den Bewertungen der einzelnen Endpunkte berücksichtigt.

#### ***Pruritus Numerical Rating Scale (Pruritus NRS)***

Der Endpunkt „Pruritus NRS“ wird zur Nutzenbewertung herangezogen und es werden die Ergebnisse zu Monat 6 dargestellt. Die Ergebnisse zu Monat 12 werden aufgrund von Unklarheiten in der Erhebung nach Monat 6 und damit einhergehend unklarer Validität ergänzend dargestellt.

#### Operationalisierung

##### *Beschreibung*

Die Pruritus NRS ist ein krankheitsspezifisches Instrument zur Messung der Intensität des Juckreizes. Eine Person beurteilt die Intensität ihres schlimmsten Juckreizes innerhalb der letzten 24 Stunden im Selbstbericht auf einer 11-stufigen Skala von „0“ (= kein Juckreiz) bis „10“ (= schlimmster vorstellbarer Juckreiz).

**Erhebung:**

- Run-in Phase bis Ende Monat 6: täglich
- Nach Monat 6: an 7 aufeinanderfolgenden Tagen pro Monat (wann genau unklar)

In der Studie RESPONSE erfolgte die Erhebung über ein elektronisches Tagebuch, das ab der Run-In-Visite bis Monat 6 täglich zu ungefähr der gleichen Zeit zwischen 16 und 20 Uhr ausgefüllt werden sollte. Der Umgang mit dem e-Tagebuch wurde in der Screening-Periode trainiert.

Nach Monat 6 wurde die Erfassung nicht mehr täglich, sondern nur noch an 7 aufeinanderfolgenden Tagen pro Monat durchgeführt. In den Studienunterlagen gibt es widersprüchliche Angaben dazu, ob und inwiefern die zeitliche Verortung der 7 Erhebungstage im Monat (bspw. jeweils zum Anfang des Monats) präspezifiziert war und angewendet wurde.

Für die Auswertung der Pruritus NRS wird zu Baseline aus den täglich erfassten Werten während der Run-in-Phase (welche  $14 \pm 3$  Tage vor Studientag 1 stattfand) und an Tag 1 der Mittelwert berechnet. Post-Baseline wurden Zieltage definiert, die den Zeitpunkten der Studienvisiten zugeordnet werden:

- nach Baseline bis Monat 6:
  - Tag 29 für Monat 1,
  - Tag 85 für Monat 3 und
  - Tag 183 für Monat 6;
- nach Monat 6 bis Monat 12:
  - Tag 242 für Monat 9 und
  - Tag 326 für Monat 12.

Für die Auswertung Post-Baseline verwendet der pU jeweils den Durchschnitt eines 7-Tage-Intervalls. Für die Zeitpunkte der Studienvisiten bis einschließlich Monat 6 werden die Pruritus-NRS-Scores der letzten 7 Tage vor dem entsprechenden Zieltag gemittelt. Für Monat 9 und 12 werden breitere Zeitfenster um den Zieltag gelegt (28 bzw. 67 Tage), aus denen 7-Tages-Daten zur Berechnung eines Durchschnittswertes berücksichtigt werden. Es ist unklar, wie bei Vorliegen von mehr als einer 7-Tages-Erhebung in einem Zeitfenster verfahren werden sollte. Die Ermittlung des 7-Tage-Durchschnitts erfolgt bereits ab Vorliegen von mindestens einem Wert.

In Modul 4 werden Analysen zu Monat 6 und zu Monat 12 vorgelegt. Im Studienprotokoll und SAP werden sowohl kontinuierliche als auch Responderanalysen zur Auswertung der Pruritus NRS definiert, welche Kapitel 2.4 zu entnehmen sind.

***Bewertung***

Die Operationalisierung ist nur in Teilen nachvollziehbar.

In den Studienunterlagen findet sich keine Begründung für das unterschiedliche Erhebungsvorgehen vor und nach Monat 6 (täglich bzw. 7 aufeinanderfolgende Tage pro Monat). Bezüglich der Erhebung nach Monat 6 ist zudem unklar, ob für die jeweils 7 aufeinanderfolgenden Erhebungstage pro Monat ein fester Zeitraum für alle Patientinnen und Patienten in beiden Behandlungsarmen präspezifiziert war, ob dieser Zeitraum individuell frei gewählt werden konnte oder ob eine gezielte Auswahl stattfand – und falls ja, wie und von wem diese Auswahl getroffen wurde. Ob demnach nach Monat 6 eine Parallelität der Erhebungszeiten beider Behandlungsarme sichergestellt ist, erscheint unklar.

Hinsichtlich der Auswertung kann nicht nachvollzogen werden und wird kritisch gesehen, dass nach Monat 6 zu den Auswertungsmonaten 9 und 12 größere Zeitfenster von bis zu 67 Tagen um den jeweiligen Auswertungszieltag definiert werden. Auch dies könnte die Parallelität der Auswertungsdaten zwischen den Behandlungsarmen gefährden.

Unabhängig vom Analysezeitraum (bis Monat 6 bzw. bis Monat 12) ist kritisch zu bewerten, dass die jeweiligen 7-Tages-Mittelwerte unabhängig von der Anzahl der vorliegenden Datenpunkte berechnet werden; das heißt, selbst wenn innerhalb des jeweiligen Zeitraums nur ein einziger Wert vorliegt, wird dieser als 7-Tages-Durchschnitt verwendet. Somit ist nicht gewährleistet, dass tatsächlich für jede Visite eine Datenbasis von 7 Tagen vorliegt. Da das Zeitfenster um den Zieltag für Monat 12 insgesamt 67 Tage breit war, können in dieses Zeitfenster zudem zwei 7-tägige Erhebungen fallen. Präspezifizierte Kriterien zum Umgang mit solchen Fällen, liegen nicht vor, was eine datengesteuerte Auswertung ermöglichen könnte.

### Patientenrelevanz

Der Endpunkt wird im vorliegenden Anwendungsgebiet grundsätzlich als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen.

### Validität

Der Endpunkt wurde in früheren Nutzenbewertungsverfahren bereits geprüft und im vorliegenden Anwendungsgebiet grundsätzlich als valide eingestuft. Jedoch werden im vorliegenden Fall aufgrund der oben beschriebenen Unklarheiten in der Operationalisierung die Validität für die Analysen nach Monat 6 als unklar bewertet. Die Analysen zu Monat 12 werden entsprechend ergänzend dargestellt. Sowohl für die Analysen zu Monat 6 als auch für die Analysen zu Monat 12 ist als mögliche Einschränkung der Validität der Ergebnisse zu beachten, dass die jeweiligen 7-Tages-Mittelwerte auch dann gebildet wurden, wenn im jeweiligen Zeitraum lediglich ein einziger Wert vorlag.

### **5-D-Itch-Scale**

Der Endpunkt „5-D Itch“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Die „5-D-Itch-Scale“ (5-D Itch) ist ein Instrument zur multidimensionalen Erfassung des Juckreizes im Zeitverlauf bei Personen mit chronischem Juckreiz. Das Instrument besteht aus 5 Domänen: „Dauer“, „Grad“, „Verlauf“, „Beeinträchtigung“ und „Lokalisation“. Die Domänen „Dauer“, „Grad“ und „Verlauf“ bestehen jeweils aus 1 Item, die Domäne „Beeinträchtigung“ besteht aus 4 Items. Alle Items dieser 4 Domänen werden anhand einer 5-Punkt-Likert-Skala im Selbstbericht bewertet. Die Domäne „Lokalisation“ umfasst 16 potenzielle Stellen, die von Juckreiz betroffen sein können. Die Anzahl der betroffenen Körperteile (Summe: 0–16) wird in 5 Kategorien geclustert: 0–2 = Punktzahl von 1; 3–5 = Punktzahl von 2; 6–10 = Punktzahl von 3; 11–13 = Punktzahl von 4; 14–16 = Punktzahl von 5. Der Bezugszeitraum war 2 Wochen. Jede Domäne kann einen Wert von 1 bis 5 Punkten erreichen, die zu erreichende Gesamtpunktzahl liegt zwischen 5 und 25 Punkten. Höhere Punkte deuten auf einen stärkeren Juckreiz hin. Ergänzend wurde ein modifizierter Gesamtscore basierend auf allen Domänen außer der Verlaufs-Domäne berechnet. Diese Anpassung erfolgte, da es laut pU häufig zu Missverständnissen bei der Bewertung dieser Domäne komme. Teilnehmende, die keinen Juckreiz haben, würden oft „unverändert“ angeben, was einem Wert von 4 entspricht, und nicht „ganz verschwunden“, was einem Wert von 1 entspricht.

Die Erhebung erfolgte analog zur Pruritus NRS im Rahmen eines elektronischen Tagebuchs. Die 5-D Itch sollte ab der Run-In-Visite (welche  $14 \pm 3$  Tage vor Studientag 1 stattfand) bis einschließlich Monat 6 alle 2 Wochen ausgefüllt werden, danach einmal pro Monat. Der Baseline-Wert wurde als arithmetisches Mittel aller verfügbaren Messungen bei Screening, während der Run-in-Phase, an Tag 1 sowie bei nicht geplanten Untersuchungen vor oder an Tag 1 berechnet.

Für jeden Zeitpunkt der Studienvisite (Monat 1, 3, 6, 9 und 12) wurde ein spezifischer Zieltag definiert (Tag 15, 57, 141, 239 und 323). Um diese Zieltage herum wurden jeweils Analysefenster gelegt (entsprechend 34, 63, 91, 91 und 98 Tage). Alle Erhebungen, die innerhalb des jeweiligen Analysefensters lagen, kamen als Wert für den Zieltag dieses Monats in Frage. Wenn mehrere geplante Erhebungen innerhalb eines Fensters lagen, wurde derjenige ausgewählt, der dem Zieltag am nächsten lag; bei gleichem Abstand wurde der später erhobene Wert verwendet. Falls keine geplante Erhebung im Analysefenster lag, aber ungeplante Erhebungen vorhanden waren (siehe Kapitel 2.3.5), wurde die ungeplante Erhebung ausgewählt, die dem Zieltag am nächsten lag. Es wurden also keine Mittelwerte aus den zweiwöchentlich erhobenen Daten gebildet.

Die Auswertung sollte gemäß SAP über die Veränderung zu den prädefinierten Zeitpunkten der Studienvisiten gegenüber Baseline erfolgen (siehe Kapitel 2.4).

#### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar. Es ist unklar, warum die Daten bis Monat 6 alle 2 Wochen und nach Monat 6 monatlich erhoben wurden, für die Auswertung jedoch vorab definierte Zieltage verwendet wurden und jeweils nur ein einziger Wert für die prädefinierten Zeitpunkte der Studienvisiten einbezogen wurde, anstatt alle erhobenen Daten zu berücksichtigen.

#### Patientenrelevanz

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

#### Validität

Die 5-D Itch wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im vorliegenden Anwendungsgebiet grundsätzlich als valide eingeschätzt.

Der pU legt einen modifizierten Gesamtscore vor, der das Item „Verlauf“ ausschließt. Diesen basiert er auf der Annahme, dass es bei der Beantwortung dieses Items zu Missverständnissen kommen kann und die Antworten der Patientinnen und Patienten somit den wahren Verlauf ihres Juckreizes ggf. nicht adäquat abbilden können (s. o.). Allerdings sind Validierungsdaten zum modifizierten Gesamtscore den Unterlagen des pU nicht zu entnehmen und die Validität des modifizierten Gesamtscores ist somit unklar. Da der nicht-modifizierte Gesamtscore der 5-D Itch umfangreich validiert ist, wird dieser für die Nutzenbewertung herangezogen. Aufgrund der beschriebenen Unsicherheiten in Bezug auf die Domäne „Verlauf“, welche in den Gesamtscore einfließt, werden die Ergebnisse der Einzeldomänen zusätzlich zum Gesamtscore berücksichtigt.

#### ***Patient Global Impression of Severity (PGI-S)***

Der Endpunkt „PGI-S“ wird in der Nutzenbewertung nicht berücksichtigt aufgrund von Mehrfacherfassung. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

## Operationalisierung

### *Beschreibung*

Mithilfe des PGI-S sollte der subjektive Eindruck des Schweregrads des Juckreizes im Selbstbericht erhoben werden. Der PGI-S bestand aus einem Item („Please choose the response below that best describes the severity of your pruritus over the past week.“), anhand welchem der Gesamtschweregrad der eigenen Symptome in der vergangenen Woche bewertet werden soll. Gemäß des im Studienprotokoll dargestellten Fragebogens erfolgte dies mit Hilfe einer 4-Punkte-Skala („kein Juckreiz“; „mild“; „moderat“; „schwer“), wobei den Kategorien keine Zahlenwerte zugeordnet wurden und dementsprechend im SAP die Auswertung rein deskriptiv vorgesehen war über die Anzahl der Teilnehmenden in den Kategorien zu den prädefinierten Zeitpunkten der Studienvisiten. In Modul 4 wird hingegen die Anwendung einer Likert-Skala von 1 (keine Symptome) bis 5 (sehr starke Symptome) beschrieben sowie post-hoc eine Auswertung mittels MMRM (Mixed Effect Model for Repeated Measurements) dargestellt.

Die Erhebung erfolgte im Rahmen des elektronischen Tagebuchs zu den präspezifizierten Zeitpunkten der Studienvisiten.

### *Bewertung*

Die Operationalisierung ist nur in Teilen nachvollziehbar. Es liegen widersprüchliche Angaben in Bezug auf die Spannweite der verwendeten Likert-Skala vor. Laut Clinical Study Report (CSR) wurden für die präspezifizierten deskriptiven Auswertungen die im Studienprotokoll definierten Kategorien („kein Juckreiz“, „mild“, „moderat“ und „schwer“), also eine 4-Punkte Likert-Skala verwendet. Dabei erfolgte keine Transformation der Kategorien in Zahlenwerte. In Modul 4 wird hingegen eine Likert-Skala mit Zahlenwerten von 1 („keine Symptome“) bis 5 („sehr starke Symptome“) beschrieben, welche demnach die Auswertung gegenüber der Erhebung gemäß Studienunterlagen um eine Kategorie erweitert und eine mindestens in Teilen abweichende Kategorienbezeichnung verwendet.

## Patientenrelevanz und Validität

Die Einschätzung der Patientinnen und Patienten zum Schweregrad der jeweils eigenen erkrankungsspezifischen Symptomatik ist als patientenrelevant zu bewerten, wobei eine Erhebung per PGI-S grundsätzlich als geeignet eingeschätzt wird.

Aufgrund der Unklarheiten bezüglich der in Modul 4 beschriebenen Kategorien und Skalenspannweite der verwendeten Likert-Skala ist die Transformation der Kategorien und damit die diesbezügliche Auswertung nicht nachvollziehbar. Es werden daher ausschließlich die präspezifizierten, deskriptiven, kategorialen Auswertungen dargestellt. Darüber hinaus besteht eine weitgehende inhaltliche Überschneidung der PGI-S mit einzelnen Domänen der 5-D Itch. Für die vorliegende Nutzenbewertung wird aufgrund dieser Mehrfacherfassung angenommen, dass der PGI-S im Vergleich zur 5-D Itch keine wesentlichen zusätzlichen Informationen generiert. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

### ***Patient Global Impression of Change (PGI-C)***

Der Endpunkt „PGI-C“ wird in der Nutzenbewertung nicht berücksichtigt aufgrund unklarer Validität und Mehrfacherfassung. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

## Operationalisierung

### *Beschreibung*

Der PGI-C wurde angewandt, um den subjektiven Eindruck des Ausmaßes der Veränderung des Juckreizes zu erfassen (Selbstbericht). Der PGI-C besteht aus einem Item („Please choose the response below that best describes the overall change in your pruritus since you started

taking the study drug.“), anhand welchem die Gesamtveränderung der eigenen Symptome im Vergleich zum Start der Studie bewertet werden sollte. Dies erfolgte auf einer 7-Punkte-Skala („viel besser“; „mäßig besser“; „ein wenig besser“; „keine Veränderung“; „ein wenig schlechter“; „mäßig schlechter“, „viel schlechter“).

Die Erhebung erfolgte analog zum PGI-S im Rahmen des elektronischen Tagebuchs zu den präspezifizierten Zeitpunkten der Studienvisiten. Analog zum PGI-S wurden den Antwortkategorien gemäß des im Studienprotokoll dargestellten Fragebogens keine Zahlenwerte zugeordnet und die Auswertung sollte entsprechend deskriptiv über die Anzahl der Teilnehmenden in den Kategorien zu den Zeitpunkten der Studienvisiten erfolgen. In Modul 4 werden den Kategorien post-hoc Zahlenwerte von 1 („sehr große Verbesserung“) bis 7 („sehr große Verschlechterung“) zugeordnet und Ergebnisse eines MMRM dargestellt.

### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz und Validität

Die Einschätzung der Patientinnen und Patienten zur Veränderung der jeweils eigenen erkrankungsspezifischen Symptomatik ist grundsätzlich als patientenrelevant zu bewerten. Ein zentrales methodisches Problem bei der Erfassung der subjektiven Veränderung der Symptomatik besteht jedoch darin, dass bei der rückwirkenden Einschätzung einer Veränderung die Befragten in der Lage sein müssen, sowohl ihren gegenwärtigen Zustand als auch ihren Ausgangszustand zu quantifizieren und daraus eine mentale Subtraktion durchzuführen. Ob dies zuverlässig möglich ist, erscheint fraglich. Es gibt Hinweise darauf, dass bei Einschätzungen von Veränderungen größtenteils der gegenwärtige Gesundheitszustand erfasst wird. Dieser Effekt wird umso größer, je länger das Erinnerungsintervall ist [12,14,15]. Dies scheint für die vorliegende Erhebung relevant zu sein, da Teilnehmende insbesondere bei der letzten Visite zu Monat 12 ein Jahr zurückblicken müssen, um sich an den Beginn ihrer Studienteilnahme zu erinnern. Für die Nutzenbewertung wird in der vorliegenden Konstellation davon ausgegangen, dass durch den PGI-C im Vergleich zur 5-D Itch keine wesentlichen, zusätzlichen Informationen generiert werden. Die Validität des PGI-C wird zudem aufgrund der oben beschriebenen methodischen Problematik als unklar bewertet. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

### **Fibrose**

Der Endpunkt „Fibrose“ wird in der Nutzenbewertung nicht berücksichtigt aufgrund nicht-gegebener Patientenrelevanz.

### Operationalisierung

#### *Beschreibung*

Die Erfassung von Fibrose erfolgte in der Studie RESPONSE mittels Enhanced-Liver-Fibrosis(ELF)-Score und der Messung der Lebersteifigkeit in Kilopascal (kPa) auf Basis einer transienten Elastographie mit FibroScan®.

Bei FibroScan® handelt es sich um einen spezialisierten Ultraschall der Leber zur Messung von Fibrose und Steatose. Die Probanden wurden angewiesen, mindestens 4 Stunden vor der geplanten Untersuchung zu fasten, sofern möglich. Die Einnahme notwendiger Medikamente und kleiner Mengen Wasser waren erlaubt. Die Scans wurden lokal überprüft und analysiert.

Der ELF-Score setzt sich aus 3 Biomarkern zusammen: Hyaluronsäure (HA), Aminotermiales Propeptid von Pro-kollagen des Typs III (PIIINP) und Gewebeinhibitor der Metalloproteinase 1 (TIMP-1). Die Erhebung dieser Biomarker erfolgte mittels Laboruntersuchungen.

Die Auswertung erfolgte über die absolute und relative Veränderung der Lebersteifigkeit gemessen anhand FibroScan® in kPa sowie des ELF-Scores zu den prädefinierten Zeitpunkten der Studienvisiten gegenüber Baseline.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Der Endpunkt erfasst ausschließlich Befunde auf Basis von Laborparametern und einem bildgebenden Verfahren ohne direkten Symptombezug, die nicht als unmittelbar patientenrelevant beurteilt werden.

Der Endpunkt „Fibrose“ wird daher in der vorliegenden Operationalisierung nicht als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen.

#### Validität

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

#### **Klinische PBC-Ereignisse**

Folgende Endpunkte wurden als klinische PBC-Ereignisse definiert:

- Lebertransplantation
- Model-for-End-Stage-Liver-Disease(MELD)-Score  $\geq 15$  bei mindestens 2 aufeinanderfolgenden Visiten
- Behandlungsbedürftiger Aszites
- Hospitalisierungen aufgrund definierter Ereignisse
- Todesfälle

Der Endpunkt „Todesfälle“ wird in der Nutzenbewertung bereits im Rahmen der „Mortalität“ berücksichtigt und in diesem Kapitel nicht erneut beschrieben.

Alle klinischen PBC-Ergebnisse wurden sowohl vom Prüfpersonal beurteilt als auch von einem CERC (Clinical Event Review Committee), bestehend aus drei unabhängigen Hepatologen und/oder Gastroenterologen. Die CERC-Mitglieder bewerteten im Rahmen einer verblindeten, unabhängigen Abstimmung, ob die Definitionen für die klinischen PBC-Ereignisse erfüllt waren. Es erfolgte eine Einordnung in die Kategorien „positiv“ (also Ereignis liegt vor), „negativ“ (Ereignis liegt nicht vor) oder „unbestimmt“.

Sobald eines der Ereignisse eintrat (durch das CERC anerkannt), wurde die entsprechende Person aus der Studie ausgeschlossen und absolvierte eine vorzeitige Abschlussvisite (siehe auch Tabelle 2 und Tabelle 6).

Für die klinischen PBC-Ereignisse erfolgt im Gegensatz zu Modul 4 im Rahmen der Nutzenbewertung eine Darstellung sowohl aller an das CERC gemeldeten Ereignisse als auch derjenigen klinischen PBC-Ereignisse, die vom CERC als bestätigt eingestuft wurden. Zudem werden die Ereignisse gemäß prüfärztlicher Einschätzung dargestellt.

#### **Lebertransplantation**

Der Endpunkt „Lebertransplantation“ wird in der Nutzenbewertung nicht berücksichtigt, aufgrund unklarer Validität, die sich aus einer unklaren Operationalisierung ergibt. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

#### Operationalisierung

##### *Beschreibung*

Zum Erhebungsvorgehen des Endpunktes konnten keine Informationen identifiziert werden.

### *Bewertung*

Die Operationalisierung ist nicht ausreichend beschrieben. Es ist u. a. unklar, wie und durch wen das Eintreten des Ereignisses ggf. systematisch erfasst wurde und wie das Ereignis definiert war. Bspw. ist nicht ersichtlich, ob auch das Setzen auf eine Warteliste als Ereignis gezählt wurde oder nur eine bereits erhaltene und erfolgreiche Lebertransplantation.

### Patientenrelevanz und Validität

Bei dem Ereignis „Lebertransplantation“ kann es sich in der vorliegenden Indikation grundsätzlich um einen patientenrelevanten Endpunkt handeln. Auch das Setzen auf eine Warteliste für eine Lebertransplantation würde als patientenrelevante Operationalisierungskomponente eingeschätzt werden. Jedoch wird aufgrund der oben beschriebenen wesentlichen Unklarheiten in Bezug auf die Operationalisierung die Validität des Endpunktes als unklar eingestuft. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

### **MELD-Score $\geq 15$ bei mindestens 2 aufeinanderfolgenden Visiten**

Der Endpunkt „MELD-Score  $\geq 15$  bei mindestens 2 aufeinanderfolgenden Visiten“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Der MELD(i)-Score (Model for End-Stage Liver Disease, initial) basiert auf den folgenden Laborparametern:

- Serumkreatinin (mg/dL)
- Gesamtbilirubin (mg/dL)
- INR (International Normalized Ratio)

Die initiale Berechnung erfolgte anhand der Formel:

$$\text{MELD}(i) = 10 \times [0,957 \times \ln(\text{Kreatinin [mg/dL]}) + 0,378 \times \ln(\text{Bilirubin [mg/dL]}) + 1,120 \times \ln(\text{INR}) + 0,643]$$

Wenn der nach dieser Formel errechnete Initialwert MELD(i) einen Wert von 11 überschreitet, wird anhand folgender Formel der MELD-Natrium(Na)-Score berechnet:

$$\text{MELD} = \text{MELD}(i) + [1,32 \times (137 - \text{Na})] - [0,033 \times \text{MELD}(i) \times (137 - \text{Na})]$$

Die Entnahme der Blutproben erfolgte analog zum Endpunkt „Biochemisches Ansprechen“.

Bei Patientinnen und Patienten mit einem MELD-Score  $\geq 15$  mussten innerhalb von 7 Tagen erneut Serumkreatinin, Gesamtbilirubin, INR und Natrium überprüft und der MELD-Score neu berechnet werden, um die Befunde zu bestätigen, eine Dekompensation zu bewerten und über einen Therapieabbruch zu entscheiden.

### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

Der Score des „Model for End-Stage Liver Disease“ (MELD) wird zur Einschätzung der Krankheitsschwere berechnet und umfasst die Parameter „Kreatinin“, „Gesamtbilirubin“, „International Normalized Ratio“ und „Serumnatriumspiegel“. Die einzelnen Bestandteile des MELD-Scores sind Laborwerte und somit ist auch der errechnete Gesamtscore nicht per se patientenrelevant. Ein MELD-Score  $\geq 15$  wird jedoch vorliegend als patientenrelevant erachtet, da dieser im deutschen Versorgungskontext die Notwendigkeit einer Lebertransplantation indiziert, ungeachtet davon, ob ein entsprechendes Spenderorgan zur Verfügung steht [1].

### Validität

Die Erhebung des Endpunkts wird als valide eingeschätzt.

#### **Behandlungsbedürftiger Aszites**

Der Endpunkt „behandlungsbedürftiger Aszites“ wird in der Nutzenbewertung nicht berücksichtigt aufgrund unklarer Validität, die sich aus einer nicht nachvollziehbaren Operationalisierung ergibt. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

### Operationalisierung

#### *Beschreibung*

Die Untersuchung hinsichtlich eines Aszites erfolgte im Rahmen der körperlichen Untersuchung zu den prädefinierten Erhebungszeitpunkten.

Weitere Informationen zum Erhebungsvorgehen konnten nicht identifiziert werden.

#### *Bewertung*

Die Operationalisierung ist nicht ausschreibend beschrieben. Es wird u. a. nicht spezifiziert, wie und durch wen der Aszites festgestellt werden sollte und anhand welcher Hilfsmittel und Kriterien die Einstufung als „behandlungsbedürftig“ erfolgte.

### Patientenrelevanz und Validität

Bei dem Ereignis „behandlungsbedürftiger Aszites“ kann es sich in der vorliegenden Indikation grundsätzlich um einen patientenrelevanten Endpunkt handeln. Jedoch wird aufgrund der oben beschriebenen Unklarheiten in Bezug auf die Erhebung die Validität des Endpunktes als unklar eingestuft. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

#### **Hospitalisierung**

Der Endpunkt „Hospitalisierung“ wird in der Nutzenbewertung nicht berücksichtigt aufgrund unklarer Validität, die sich aus einer nicht nachvollziehbaren Operationalisierung ergibt. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

### Operationalisierung

#### *Beschreibung*

Erfasst wurden Hospitalisierungen aufgrund eines neuen Auftretens oder eines Wiederauftretens von:

- Varizenblutung;
- hepatischer Enzephalopathie (definiert durch einen West-Haven-Score  $\geq 2$ );
- spontaner bakterieller Peritonitis (bestätigt durch Kultur aus diagnostischer Parazentese).

Weitere Informationen zum Erhebungsvorgehen konnten nicht identifiziert werden.

Die Untersuchung hinsichtlich einer hepatischen Enzephalopathie erfolgte im Rahmen der körperlichen Untersuchung zu den prädefinierten Erhebungszeitpunkten.

#### *Bewertung*

Die Operationalisierung ist nicht ausreichend beschrieben. Den Studienunterlagen konnte u. a. nicht entnommen werden, wie und anhand welcher Hilfsmittel der West-Haven-Score erhoben bzw. ermittelt wurde. Ebenso ist unklar, wie und durch wen die Feststellung einer Varizenblutung oder einer spontanen bakteriellen Peritonitis erfolgte. Eine Definition, welche Länge und Art der Hospitalisierung (aufgrund der genannten Symptome) im Rahmen des Endpunktes gewertet wurde und welche nicht, liegt ebenfalls nicht vor.

### Patientenrelevanz und Validität

Bei dem Ereignis „Hospitalisierung“ kann es sich in der vorliegenden Indikation grundsätzlich um einen patientenrelevanten Endpunkt handeln. Jedoch wird aufgrund der oben beschriebenen Unklarheiten in Bezug auf die Erhebung die Validität des Endpunktes als unklar eingestuft. Aus Transparenzgründen erfolgt eine Darstellung im Anhang.

### **2.3.3 Lebensqualität**

#### **PBC-40**

Der Endpunkt „PBC-40“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der Fragebogen „PBC-40“ ist ein krankheitsspezifisches Instrument zur Messung der Lebensqualität bei Personen mit PBC. Der patientenberichtete Fragebogen besteht aus 40 Items in 6 Domänen und wird anhand einer 5-stufigen Skala (von „nie“/„stimme überhaupt nicht zu“ bis „immer“/„stimme voll und ganz zu“) beantwortet:

- Generelle Symptome (7 Items)
- Juckreiz (3 Items)
- Fatigue (11 Items)
- Kognitive Funktion (6 Items)
- Soziale Funktion (10 Items)
- Emotionale Funktion (3 Items)

Der Bezugszeitraum des PBC-40 beträgt 4 Wochen. Die Erhebung erfolgte über das elektronische Tagebuch zu den prädefinierten Zeitpunkten der Studienvisiten.

Zur Auswertung werden den Item-Antworten laut Modul 4 sogenannte Score-Punkte zugeordnet (von 1 = „nie“/„stimme überhaupt nicht zu“ bis 5 = „immer“/„stimme voll und ganz zu“). Die Score-Punkte können dann zu den jeweiligen Domänen-Scores aufsummiert werden, wobei ein Domänenscore gemäß Manual [13] nur bei einer Beantwortung von mindestens 50 % der Items vorgesehen ist. Höhere Werte spiegeln dabei eine geringere Lebensqualität wider. Der Gesamt-Score ergibt sich laut Modul 4 aus der Aufsummierung aller Items und kann somit Werte zwischen 32 und 200 annehmen.

Die Auswertung erfolgte laut SAP über die Veränderung zu den präspezifizierten Zeitpunkten der Studienvisiten gegenüber Baseline.

##### *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

Der Endpunkt „PBC-40“ wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im vorliegenden Anwendungsgebiet als valide eingeschätzt. Ein domänenübergreifender Gesamtscore ist für den PBC-40 nicht validiert [11]. In der Nutzenbewertung erfolgt daher die Darstellung der Ergebnisse für die einzelnen Domänen.

## 2.3.4 Sicherheit

### Unerwünschte Ereignisse

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

#### Operationalisierung

##### *Beschreibung*

Unerwünschte Ereignisse (UE) sind definiert als jegliche ungünstigen und ungewollten medizinischen Zwischenfälle oder Ereignisse, die bei einer Patientin oder einem Patienten auftreten, während diese oder dieser ein Prüfpräparat erhält. Dazu zählen auch anormale Laborbefunde, Symptome oder jegliche Krankheiten, die zeitlich mit der Anwendung eines Arzneimittels zusammenhängen – unabhängig davon, ob ein kausaler Zusammenhang mit der Studienbehandlung besteht. Ebenso umfasst die Definition Zustände, die durch Begleiterkrankungen, Reaktionen auf Begleitmedikamente oder das Fortschreiten von Krankheitszuständen verursacht werden.

Alle UE werden ab dem Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis zum Ende der Studienteilnahme (letzte Studienvisite) kontinuierlich erhoben. Teilnehmende, die die Behandlung mit dem Studienmedikament abbrechen, wurden gebeten, ohne Einnahme des Studienmedikaments weiter an der Studie teilzunehmen (außer Teilnehmende, die aufgrund eines PBC-Ereignisses die Studienmedikation abbrechen).

Abnormale Laborbefunde werden durch Überprüfung aller bei den Teilnehmenden erhobenen Labordaten festgestellt. Jede Laboranomalie, die von der Prüffärztin/dem Prüffarzt als klinisch bedeutsam erachtet wird, sollte als UE gemeldet werden. Eine klinisch bedeutsame Anomalie ist eine bestätigte Anomalie, die gegenüber Baseline ausreichend verändert ist, sodass nach dem klinischen Urteil der Prüffärztin/des Prüffarztes eine Änderung des weiteren Vorgehens gerechtfertigt ist. Diese Änderung konnte ein verstärktes Monitoring des Laborwerts, die Einleitung weiterer diagnostischer Tests oder Verfahren sowie die Verabreichung einer Behandlung umfassen.

Es werden alle UE als TEAE (Treatment Emergent Adverse Event) definiert, die nach Beginn der Behandlung mit dem Studienmedikament bis zum Ende der Sicherheitsnachbeobachtung nach der letzten Verabreichung des Studienmedikaments neu aufgetreten sind, in ihrer Häufigkeit zugenommen oder sich in ihrem Schweregrad verschlechtert haben. Mit dem finalen Studienprotokoll Version 4 vom 9. Februar 2022 (etwa 10 Monate nach Einschluss der ersten Studienperson) wurde die Sicherheitsnachbeobachtung von einem Monat auf 14 (+3) Tage gekürzt. Laut Studienprotokoll entfiel zudem bei einem Wechsel in die Langzeitnachbeobachtungsstudie ASSURE nach Monat 12 die reguläre Sicherheitsnachbeobachtung der RESPONSE-Studie. Die Auswertung sollte gemäß SAP jedoch für den gesamten Zeitraum bis 30 Tage nach der letzten Verabreichung der Studienmedikation erfolgen.

Die UE wurden gemäß „Medical Dictionary for Regulatory Activities“ (MedDRA), Version 26.0, nach Systemorganklasse (SOC) und Preferred Term (PT) kodiert. Die Einteilung in Schweregrade erfolgte gemäß „Common Terminology Criteria for Adverse Events“ (CTCAE), Version 5.0.

Als schwerwiegende unerwünschte Ereignisse (SUE) werden UE definiert, die bei jeglicher Dosis:

- zum Tode führten;
- direkt lebensgefährdend waren;
- eine Hospitalisierung bzw. Verlängerung einer Hospitalisierung erforderten (keine elektiven Aufnahmen);

- zu einer anhaltenden oder erheblichen Behinderung oder Beeinträchtigung führten (d. h. zu einer erheblichen Beeinträchtigung der Fähigkeit einer Person, normale Lebensfunktionen auszuüben);
- eine kongenitale Anomalie / angeborene Fehlbildung darstellten oder
- als medizinisch bedeutsames Ereignis eingestuft wurden.
  - = medizinisches Ereignis, das nach angemessenem medizinischem Ermessen die Person gefährden kann und möglicherweise einen medizinischen oder chirurgischen Eingriff erfordert, um eines der oben in der Definition für ein SAE aufgeführten Ergebnisse zu verhindern.

Todesfälle, die nach dem letzten Studienbesuch eintreten, werden nicht als SUE gemeldet.

Es werden zudem UE von besonderem Interesse (AESI) definiert. Diese umfassen Pruritus-assoziierte TEAE sowie TEAE, die potenziell auf folgende Toxizitäten hinweisen:

- Lebertoxizität (SMQ (Standardised MedDRA Queries) „Erkrankungen der Leber“)
- Muskelttoxizität (breite FMQ (FDA MedDRA Queries) „Myalgie“)
- Nierentoxizität (SMQ „Akutes Nierenversagen“)
- Pankreasbezogene Toxizität (breite FMQ „Pankreatitis“)
- Kardiovaskuläre Toxizität
  - Herzrhythmusstörungen (SMQ)
  - Herzinsuffizienz (SMQ)
  - Kardiomyopathie (SMQ)
  - Ischämische Herzerkrankung (SMQ)

In Modul 4 werden zudem – im Gegensatz zu den Studienunterlagen – Frakturen als AESI gezählt. Da die Einordnung von Frakturen als AESI den Studienunterlagen nicht entnommen werden konnte, wird diese als post-hoc eingestuft.

### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar.

Gemäß Studienprotokoll entfiel bei einem Wechsel in die Langzeitnachbeobachtungsstudie ASSURE nach Monat 12 die reguläre Sicherheitsnachbeobachtung der RESPONSE-Studie. Es ist also zu berücksichtigen, dass für Teilnehmende, die in die ASSURE-Studie wechselten, keine Daten für die Sicherheitsnachbeobachtung vorliegen.

Zudem ist der Umgang hinsichtlich der Erfassung von Progressionen von Krankheitszuständen als UE im Protokoll nicht eindeutig formuliert. Das im Studienprotokoll verwendete Wording lässt offen, ob lediglich die Progression eines Krankheitszustandes, der als vorbestehendes UE eingestuft wurde, als UE gezählt wird, oder ob möglicherweise auch eine Progression der Grunderkrankung als UE dokumentiert wurde.

### Patientenrelevanz

Der Endpunkt „Unerwünschte Ereignisse“ wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen. Die Patientenrelevanz von Laborparametern ist unklar.

### Validität

Die Erhebung des Endpunktes wird als weitgehend valide erachtet.

Es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung mit in die Erfassung der UE eingingen und die Ergebnisse zu den UE somit teilweise auch Wirksamkeitsaspekte abbilden könnten. Beispielsweise sind „Pruritus-assoziierte TEAE“ als AESI definiert und im Studienprotokoll ist – wie in der Bewertung der Operationalisierung ausgeführt – nicht eindeutig ersichtlich, ob auch Progressionen der Grunderkrankung als UE erfasst wurden. Der

pU legt in Modul 4 zwar für die Gesamtraten zusätzliche Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor, da aber nicht ersichtlich ist, welche erkrankungsbezogenen Ereignisse dabei nicht berücksichtigt wurden, werden diese Analysen nicht dargestellt. Für die UE nach SOC und PT liegen keine entsprechenden Auswertungen vor.

### 2.3.5 Erhebungszeitpunkte

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

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

Studienvisite Endpunkt	Screening <sup>1)</sup>	Run- In	R.	M1	M3	M6	M9	M12/ EOT	FU <sup>2)</sup>	ET <sup>3)</sup>	UNS <sup>4)</sup>
	T-35 bis -15	T-14 ± 3	T1	T29 ± 3	T85 ± 7	T183 ± 7	T274 ± 7	T365 ± 7	14T ± 3		
Todesfälle <sup>5)</sup>	kontinuierlich										
Pruritus											
Pruritus NRS	x	x <sup>6)</sup>	x <sup>7)</sup>	x <sup>7)</sup>		x <sup>7)</sup>	(x) <sup>4)</sup>				
5-D Itch		x <sup>8)</sup>	x <sup>9)</sup>	x <sup>9)</sup>		x <sup>9)</sup>	(x) <sup>4)</sup>				
MELD-Score				x	x	x	x	x	x	x	
PBC-40		x	x	x	x	x	x	x		x	(x) <sup>4)</sup>
Unerwünschte Ereignisse	kontinuierlich										

<sup>1)</sup> Mit Protokoll-Amendment 4 (09.02.2022) wurde das Screening-Zeitfenster von bis zu 2 auf bis zu 3 Wochen verlängert.

<sup>2)</sup> Mit Protokoll-Amendment 4 (09.02.2022) wurde die Dauer der Sicherheitsnachbeobachtung auf bis zu 17 Tage nach der letzten Gabe des Studienmedikaments festgelegt. Frühere Protokollversionen sahen hierfür einen Zeitraum von einem Monat vor. Für Teilnehmende, die nach Monat 12 in die Langzeitnachbeobachtungsstudie ASSURE wechselten, entfiel die reguläre Sicherheitsnachbeobachtung im Rahmen der RESPONSE-Studie.

<sup>3)</sup> Teilnehmende, die die Behandlung mit dem Studienmedikament aus anderen Gründen als einem klinischen PBC-Ereignis beendeten, wurden gebeten, weiterhin an der Studie ohne Einnahme des Studienmedikaments teilzunehmen. Wurde dies abgelehnt und die Studienteilnahme zu einem beliebigen Zeitpunkt nach Tag 1 beendet, wurde eine Early-Termination(ET)-Visite durchgeführt.

<sup>4)</sup> Ein „Unscheduled (UNS-) Visit“, also eine ungeplante Visite, musste bei Erfüllung präspezifizierter Sicherheitskriterien oder bei Dosisanpassung erfolgen und konnte jederzeit nach ärztlicher Einschätzung erfolgen, um die Sicherheit der Teilnehmenden zu gewährleisten. Pruritus NRS, 5-D Itch, PBC-40, PGI-S und PGI-C waren im elektronischen Tagebuch für ungeplante Besuche verfügbar, wenn dies als notwendig erachtet wurde.

<sup>5)</sup> Im Rahmen der Sicherheitsüberwachung erhoben.

<sup>6)</sup> Tägliche Erhebung über ein elektronisches Tagebuch.

<sup>7)</sup> Erhebung an jeweils 7 aufeinanderfolgenden Tagen pro Monat (welche Tage unklar; siehe Kapitel 2.3) über ein elektronisches Tagebuch.

<sup>8)</sup> Erhebung alle 2 Wochen im Rahmen des elektronischen Tagebuchs.

<sup>9)</sup> Nach Monat 6 erfolgte die Erhebung einmal pro Monat im Rahmen des elektronischen Tagebuchs.

Abkürzungen: EOT: Ende der Behandlung; ET: Early Termination; FU: Follow-Up; MELD: Model for End-Stage Liver Disease; M: Monat; NRS: Numerical Rating Scale; PBC: Primär biliäre Cholangitis; PGI-C: Patient Global Impression of Change; PGI-S: Patient Global Impression of Severity; R.: Randomisierung; T: Tag; UNS: Unscheduled.

## 2.4 Statistische Methoden

Die statistischen Methoden basieren auf dem SAP der Studie RESPONSE, der auf den 28.08.2023 datiert ist, sowie auf dem ergänzenden SAP vom 29.08.2023, welcher zusätzliche explorative Analysen spezifiziert. Beide SAP-Dokumente wurden nach dem für die Nutzenbewertung relevanten Datenschnitt am 11.08.2023 und kurz vor Datenbankschließung am 30.08.2023 erstellt.

### Analysepopulationen

- Intention-to-Treat(ITT)-Population: Alle Personen, die in die Studie randomisiert wurden und mindestens eine Dosis des Studienmedikaments erhalten haben. Teilnehmende werden entsprechend der randomisierten Behandlungszuweisung analysiert.
- Moderate-to-Severe-Pruritus-NRS(MSPN)-Population: Alle Personen aus der ITT-Population, die einen NRS-Score  $\geq 4$  zu Baseline haben.
- Sicherheitspopulation: Alle Personen, die mindestens eine Dosis des Studienmedikaments erhalten haben. Teilnehmende werden in die Gruppe aufgenommen, die der tatsächlich erhaltenen Behandlung entspricht, sofern diese vom ursprünglichen Behandlungsarm abweicht.

Die Wirksamkeitsanalysen wurden i. d. R. in der ITT-Population durchgeführt. Für den Endpunkt „Pruritus NRS“ waren auch Auswertungen in der MSPN-Population präspezifiziert.

Zur Nutzenbewertung wird die ITT- und die Sicherheitspopulation herangezogen, im vorliegenden Fall stimmen diese überein.

### Datenschnitte

Der Zeitpunkt des Datenschnitts geht nicht eindeutig aus den Studienunterlagen hervor.

In Modul 4 wird der 11.08.2023 als Datenschnitt angegeben. Zwar finden sich im Studienbericht keine eindeutigen Angaben zum Datenschnitt, jedoch ist der 11.08.2023 dort als Zeitpunkt der letzten Visite der letzten Studienperson dokumentiert. Es waren keine Zwischenanalysen vorgesehen; die Analyse der Daten sollte gemäß SAP nach Abschluss der letzten Visite aller Studienteilnehmenden und nach finaler Datenbanksperre durchgeführt werden.

### Präspezifizierte Subgruppenanalysen

Es wurden u. a. folgende Subgruppenmerkmale für Subgruppenanalysen präspezifiziert:

- Alter beim Screening ( $< 65$ ,  $\geq 65$  Jahre)
- Geschlecht (Weiblich, Männlich)
- Region (Nordamerika, Europa, Restliche Welt)
- Baseline Gesamtbilirubin ( $< 0,6 \times \text{ULN}$ ,  $\geq 0,6 \times \text{ULN}$ )
- Baseline Pruritus NRS ( $< 4$ ,  $\geq 4$ )
- Baseline Gesamtbilirubin ( $\leq 1 \times \text{ULN}$ ,  $> 1 \times \text{ULN}$ )

Beim Vorliegen signifikanter Interaktionsterme erfolgt für in der Nutzenbewertung berücksichtigte Endpunkte eine Angabe der Ergebnisse der Subgruppenanalysen.

### Präspezifizierung geplanter und durchgeführter Analysen

Für die 5-D Itch waren gemäß Studienunterlagen deskriptive Auswertungen der Veränderung gegenüber Baseline vorgesehen. In Modul 4 und im Studienbericht werden kontinuierliche Analysen mittels MMRM dargestellt, die als post-hoc eingestuft werden.

Auch die Auswertung der klinischen PBC-Ereignisse sollte gemäß SAP rein deskriptiv über die Inzidenz (gemäß Beurteilung des Prüfarztes/der Prüfarztin sowie gemäß der Beurteilung durch das CERC) stattfinden.

Die Auswertung des Endpunktes „PBC-40“ erfolgte, wie im SAP präspezifiziert, kontinuierlich über die Veränderung zu den präspezifizierten Zeitpunkten der Studienvisiten gegenüber Baseline mittels MMRM.

### Pruritus NRS

Im Studienprotokoll und SAP werden für die Pruritus NRS zahlreiche stetige Analysen, u. a. mittels MMRM, sowie Responderanalysen definiert. Präspezifiziert in der ITT-Population waren u. a.:

- MMRM:
  - Veränderung der Pruritus-NRS-Scores zu Monat 1, 3, 6, 9 und 12 gegenüber Baseline auf Basis der gemittelten 7-Tage-Intervalle.
  - Veränderung des Maximums der Pruritus-NRS-Scores für Monat 1, 3 und 6 gegenüber Baseline auf Basis der 7-Tage-Intervalle.
- Responderanalysen:
  - Anteil der Teilnehmenden mit Pruritus NRS > 0 zu Baseline und einer Reduktion um > 4 Punkte oder mit einer Reduktion auf Pruritus NRS = 0 auf Basis der gemittelten 7-Tage-Intervalle.
  - Anteil der Teilnehmenden mit Pruritus NRS > 0 zu Baseline und mit einer Reduktion auf Pruritus NRS = 0 auf Basis der gemittelten 7-Tage-Intervalle.
- Fläche unter der Kurve (AUC) für tägliche Änderungen der Pruritus NRS gegenüber Baseline (AUC<sub>Δ</sub>).

In Modul 4 wird eine Responderanalyse für die Anzahl und den Anteil an Personen mit einer Verringerung um  $\geq 4$  Punkte gegenüber Baseline dargestellt sowie die stetige Analyse der Veränderung gegenüber Baseline zu den Monaten 6 und 12 mittels MMRM unter Verwendung der gemittelten 7-Tage-Intervalle in der ITT. Die Responderanalyse wird als post-hoc eingestuft.

Alle präspezifizierten Analysen sollten nach Möglichkeit für die Stratifizierungsfaktoren (Pruritus-NRS- und ALP-Werte zu Baseline) adjustiert werden.

### Responderanalysen

In Modul 4 wird post-hoc das Relative Risiko (RR) mit zugehörigem p-Wert anhand der Mantel-Haenszel-Methode berichtet. Die 95%-Konfidenzintervalle und p-Werte wurden unter Verwendung der Normalapproximation (Wald-Test) berechnet. Auch wenn die Mantel-Haenszel-Methode eingesetzt werden kann, um einen gepoolten Schätzer über mehrere Strata hinweg zu berechnen, bleibt im vorliegenden Fall unklar, ob das RR tatsächlich für jeden Endpunkt stratifiziert nach den Randomisierungsfaktoren berechnet wurde.

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

### MMRM

Fehlt im Rahmen der Pruritus-NRS-Analyse mittels MMRM ein Messzeitpunkt, wird dieser als Mittelwert der beiden benachbarten wöchentlichen Durchschnittswerte (maximal 1 Woche Abstand) imputiert; ist nur ein benachbarter Wert vorhanden, wird dieser zur Imputation verwendet. Fehlen bei einer Person bspw. die Daten für Woche 23 und 24, wird Woche 23 anhand von Woche 22 und Woche 24 anhand der Daten von Woche 25 imputiert. Bricht eine Person die Studie vor oder während Woche 24 ab, wird für Woche 26 kein Wert imputiert. Für

die monatlich gemittelten Werte der Pruritus-NRS-Analysen werden fehlende Monate durch den Mittelwert der beiden benachbarten Monate ersetzt (falls nur einer existiert, wird dieser übernommen). Daten, die nach Monat 6 erhoben wurden, werden für die Imputation nicht verwendet.

### Responderanalysen

Jede Person, für die keine Bewertung vorliegt, die die Behandlung vor dem festgelegten Zeitpunkt der Response-Bewertung abgebrochen hat oder aus anderen Gründen fehlende Daten aufweist, wird als Nicht-Responder gewertet.

### **Einschätzung der statistischen Auswertungen**

Für die Nutzenbewertung werden die Ergebnisse der ITT-Population herangezogen, da diese die gesamte zulassungskonforme Population abbildet.

In Modul 4 werden Analysen innerhalb eines Full-Analysis-Sets (FAS) dargestellt, ein FAS wurde in den Studienunterlagen jedoch nicht definiert. Die FAS-Analysen in Modul 4 entsprechen den Ergebnissen der ITT-Analysen aus dem Studienbericht.

Bei dem vorliegenden Anwendungsgebiet handelt es sich um eine langsam progrediente Erkrankung, bei der die Auswertung unter Berücksichtigung eines längeren Zeitverlaufs gegenüber einer Auswertung zu nur einem Erhebungszeitpunkt bevorzugt wird. Daher werden die Auswertungen der kontinuierlichen Daten mittels MMRM bevorzugt. Zudem wird grundsätzlich der späteste Beobachtungszeitpunkt, also Monat 12, für die Bewertung bevorzugt. Soweit vorhanden, werden zusätzlich Responderanalysen dargestellt.

Für die kontinuierliche Auswertung der Pruritus NRS wurden MMRM-Analysen durchgeführt. Zwar werden grundsätzlich die Ergebnisse zum spätesten Beobachtungszeitpunkt bevorzugt, aufgrund der in Kapitel 2.3.2 beschriebenen Unklarheiten im Erhebungsvorgehen nach Monat 6 und der damit verbundenen unklaren Validität, werden die Analysen zu Monat 12 jedoch lediglich ergänzend dargestellt. Für die Nutzenbewertung werden primär die Analysen zu Monat 6 herangezogen.

Die prädefinierte Responderanalyse, die den Anteil der Teilnehmenden mit einem Pruritus-NRS-Wert  $> 0$  zu Baseline und lediglich einer Reduktion auf Pruritus NRS = 0 betrachtet, berücksichtigt jene Personen nicht als Responder, die eine relevante Verbesserung auf der Pruritus NRS zeigen, jedoch mit hohen Ausgangswerten starten und im Verlauf der Studie keinen Wert von 0 erreichen. Diese Teilnehmenden würden trotz relevanter Symptomlinderung nicht als Responder klassifiziert werden. Die in Modul 4 dargestellte post-hoc Responderanalyse mit einer Reduktion um mindestens 4 Punkte ist hingegen ausschließlich für Teilnehmende gültig, die zu Baseline einen Wert von mindestens 4 Punkten aufwiesen. Das Erreichen einer kompletten Reduktion des Juckreizes wird vorliegend jedoch auch als relevant erachtet. Daher wird die präspezifizierte kombinierte Responderanalyse für Personen mit einem Baseline-Wert  $> 0$  und einer Reduktion um 4 Punkte oder auf 0 dargestellt. Analysen, die alle Teilnehmenden der ITT-Population einbeziehen und Personen mit einem Baseline-Wert von 0 als Non-Responder klassifizieren, würden grundsätzlich bevorzugt werden. Solche Auswertungen liegen jedoch nicht vor. Die prozentualen Angaben der eingereichten Analysen werden daher im Ergebnisteil in Eigenberechnung an die ITT angepasst. Analog zu den MMRM-Analysen werden primär die Ergebnisse zu Monat 6 dargestellt und die Ergebnisse zu Monat 12 ergänzend.

Für die 5-D Itch, und den PBC-40 werden die stetigen Analysen zur Veränderung gegenüber Baseline zu Monat 12 mittels MMRM dargestellt.

Für den Endpunkt „MELD-Score  $\geq 15$  bei mindestens 2 aufeinanderfolgenden Visiten“ werden die präspezifizierten deskriptiven Inzidenzen herangezogen.

## 2.5 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 7: Verzerrungspotential der Studie RESPONSE

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
RESPONSE	Ja <sup>1)2)</sup>	Ja	Ja <sup>3)</sup>	Ja	Nein	Nein	Niedrig

<sup>1)</sup> Die Randomisierung erfolgte zentral über ein interaktives webbasiertes Antwortsystem (IWRS) zur Visite an Studientag 1. Der Randomisierungsplan wurde von einem nicht verblindeten Statistiker vorbereitet, der nicht zum Studienteam gehörte.

<sup>2)</sup> Insgesamt wurden laut Studienbericht 11 Personen einem falschen Stratum zugeordnet, davon 10 in der Seladelpar-Gruppe (7,8 %) und 1 in der Placebo-Gruppe (1,5 %). Es bleibt jedoch unklar, wie viele dieser Probanden fälschlicherweise dem Stratum mit ALP  $< 350$  U/l anstelle des Stratum mit ALP  $\geq 350$  U/l zugeordnet wurden bzw. wie viele anstelle des Stratum mit Pruritus NRS  $\geq 4$  im Stratum mit Pruritus NRS  $< 4$  eingeordnet wurden. Die Baseline-Charakteristika der ITT-Population lassen jedoch darauf schließen, dass die Randomisierung insgesamt adäquat durchgeführt wurde. Laut EPAR hatten die fehlerhaften Zuordnungen keinen Einfluss auf die Studienergebnisse.

<sup>3)</sup> Die Art der Verabreichung, das Aussehen und das Einnahmeschema von Seladelpar und Placebo waren identisch. Abkürzungen: ALP: Alkalische Phosphatase; EPAR: European Public Assessment Report; ITT: Intention to Treat; NRS: Numerical Rating Scale.

Das Verzerrungspotential auf Studienebene wird als niedrig eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 8 dargestellt.

Tabelle 8: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie RESPONSE

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
<b>Morbidität</b>					
Pruritus					
Pruritus NRS					
• Pruritus NRS zu Monat 6	Ja	Unklar <sup>1)2)</sup>	Nein	Nein	Unklar
• Pruritus NRS zu Monat 12 (ergänzend dargestellt)	Ja	Nein <sup>1)3)</sup>	Ja <sup>4)</sup>	Ja <sup>5)</sup>	Hoch
5-D Itch	Ja	Nein <sup>6)</sup>	Nein	Ja <sup>7)</sup>	Hoch

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
<b>Lebensqualität</b>					
PBC-40	Ja	Nein <sup>8)</sup>	Nein	Nein	Hoch
<b>Sicherheit</b>					
UE	Ja	Ja	Nein	Nein <sup>9)10)</sup>	Niedrig

<sup>1)</sup> Die Pruritus NRS wurde über ein e-Tagebuch erfasst, bis einschließlich Monat 6 täglich und nach Monat 6 für 7 aufeinanderfolgende Tage pro Monat. Für die Auswertung Post-Baseline verwendet der pU jeweils den Durchschnitt eines 7-Tage-Intervalls, welcher den Zeitpunkten der Studienvisiten zugeordnet wird. Aufgrund dieser Methodik lässt sich nicht nachvollziehen, wie viele der tatsächlichen Erhebungen fehlen (siehe Kapitel 2.3.2). Der Durchschnitt der 7-Tage-Intervalle wurde zudem bereits berechnet, wenn nur ein einziger Wert vorlag, sodass die Ergebnisse möglicherweise auf jeweils nur einem Wert über die jeweils 7-tägige Spanne pro Auswertungs-Zeitpunkt basieren. Daten zur Vollständigkeit der e-Tagebuch-Einträge liegen nicht vor. Dadurch kann nicht ausgeschlossen werden, dass fehlende Werte zwischen den Studienarmen ungleich verteilt waren.

<sup>2)</sup> Die Rücklaufquoten in den Studienarmen waren vergleichbar (Seladelpar: 94,5 %; Placebo: 93,8 %). Da jedoch keine Daten zur Vollständigkeit der e-Tagebuch-Einträge vorliegen und somit nicht ausgeschlossen werden kann, dass fehlende Werte zwischen den Studienarmen ungleich verteilt waren, wird von einem unklaren Bias-Risiko ausgegangen.

<sup>3)</sup> Die Rücklaufquoten zu Monat 12 lagen im Seladelpar-Arm bei 82 % und im Placebo-Arm bei 73,8 %. Zudem fehlen Angaben zur Vollständigkeit der e-Tagebuch-Einträge. Auch wenn die angegebenen Rücklaufquoten zwischen den Studienarmen weitgehend vergleichbar sind, wird von einem hohen Bias-Risiko ausgegangen.

<sup>4)</sup> Für die Auswertung Post-Baseline verwendet der pU jeweils den Durchschnitt eines 7-Tage-Intervalls, welcher den Zeitpunkten der Studienvisiten zugeordnet wird. Für Monat 1, 3 und 6 werden die Pruritus-NRS-Scores der letzten 7 Tage vor dem entsprechenden Zieltag gemittelt. Für Monat 9 und 12 werden breitere Zeitfenster um den jeweiligen Zieltag gelegt (28 bzw. 67 Tage), aus denen Erhebungen aus einem 7-Tages-Intervall zur Berechnung eines Durchschnittswertes berücksichtigt werden. Da das Zeitfenster um den Zieltag für Monat 12 insgesamt 67 Tage breit war, können in dieses Zeitfenster zwei 7-tägige Erhebungen fallen. Präspezifizierte Kriterien zum Umgang mit solchen Fällen (also bspw., dass immer das spätere 7-Tage-Intervall berücksichtigt werden sollte) liegen nicht vor, was eine datengesteuerte Auswertung ermöglichen könnte.

<sup>5)</sup> Das Erhebungsvorgehen nach Monat 6 ist unklar (siehe Kapitel 2.3.2), möglicherweise sind keine parallelen Messungen zwischen den Teilnehmenden und somit auch zwischen den Studienarmen sichergestellt.

<sup>6)</sup> Die 5-D Itch wurde im Rahmen des e-Tagebuchs bis einschließlich Monat 6 zweiwöchentlich und nach Monat 6 monatlich erfasst. Aufgrund der Methodik, wie den Zeitpunkten der Studienvisiten Werte für die 5-D Itch zugeordnet wurden, lässt sich nicht nachvollziehen, wie viele der tatsächlichen Erhebungen fehlen (siehe Kapitel 2.3.2). Daten zur Vollständigkeit der e-Tagebuch-Einträge liegen nicht vor. Dadurch kann nicht ausgeschlossen werden, dass fehlende Werte zwischen den Studienarmen ungleich verteilt waren. Die Rücklaufquoten zu Monat 12 lagen im Seladelpar-Arm bei 75 % und im Placebo-Arm bei 70,8 %. Auch wenn die angegebenen Rücklaufquoten zwischen den Studienarmen weitgehend vergleichbar sind, wird auf Grund der dargelegten Einschränkungen von einem hohen Bias-Risiko ausgegangen.

<sup>7)</sup> Für die Zuordnung der 5-D-Itch-Werte zu den jeweiligen Zeitpunkten der Studienvisiten wurden Zieltage definiert, um die teilweise sehr breite Zeitfenster gelegt wurden. Aus diesen Zeitfenstern wurde jeweils der Wert ausgewählt, der dem Zieltag für die Studienvisite am nächsten lag und bei gleichem Abstand wurde der spätere Termin verwendet (siehe Kapitel 2.3.2). Aufgrund der Breite dieser Zeitfenster könnten die gewählten Werte zeitlich weit vom Zieltag entfernt liegen. Dadurch ist die Parallelität der Erhebungen zwischen den Teilnehmenden und somit auch den Studienarmen möglicherweise nicht mehr gewährleistet.

<sup>8)</sup> Die Rücklaufquoten zu Monat 12 lagen im Seladelpar-Arm bei 73,4 % und im Placebo-Arm bei 78,5 %. Auch wenn die angegebenen Rücklaufquoten zwischen den Studienarmen weitgehend vergleichbar sind, wird von einem hohen Bias-Risiko ausgegangen.

<sup>9)</sup> Für Teilnehmende, die in die Langzeitnachbeobachtungsstudie „ASSURE“ wechselten, entfiel die reguläre Sicherheitsnachbeobachtung von  $14 \pm 3$  Tagen nach Monat 12/EOT. Da der Anteil der in die ASSURE-Studie gewechselten Teilnehmenden zwischen den Studienarmen vergleichbar war, wird nicht von einer Verzerrung ausgegangen.

<sup>10)</sup> Es ist unklar, ob adjustierte Effektschätzer vorliegen. Aufgrund der vorliegenden Ergebnisse werden jedoch keine Auswirkungen auf das Verzerrungspotential erwartet.

Abkürzungen: EOT: Ende der Behandlung; ITT: Intention to Treat; NRS: Numerical Rating Scale; PBC: Primär biliäre Cholangitis; pU: pharmazeutischer Unternehmer; UE: Unerwünschtes Ereignis.

### 3 Ergebnisse der eingeschlossenen Studie

Insgesamt wurden 360 Personen gescreent, von denen 193 in die Studie eingeschlossen wurden – 128 in den Seladelpar-Arm und 65 in den Placebo-Arm. Der Einschluss in die Studie wurde im Juli 2022 abgeschlossen. Studienübergreifend schlossen 175 Personen (91 %) die Behandlungsphase ab, davon 118 im Seladelpar-Arm (92 %) und 57 im Placebo-Arm (88 %).

Die mediane Beobachtungsdauer wird in Modul 4 post-hoc für jeden Endpunkt separat dargestellt und daher bei den Ergebnissen zu den einzelnen Endpunkten in den Kapiteln 3.2 – 3.5 entsprechend aufgeführt.

#### 3.1 Studiencharakteristika und Studienmedikation

Tabelle 9: Allgemeine Angaben; Studie RESPONSE (Ende der Studie: 11.08.2023)

Studie RESPONSE Allgemeine Angaben	Seladelpar N = 128	Placebo N = 65
ITT-Population <sup>1)</sup> , n (%)	128 (100)	65 (100)
Sicherheitspopulation <sup>1)</sup> , n (%)	128 (100)	65 (100)
Abbruch der Studienteilnahme, n (%)	11 (8,6)	8 (12,3)
UE	3 (2,3)	4 (6,2)
Klinisches PBC-Ereignis	0 <sup>2)</sup>	0
Lost to Follow-up	1 (0,8)	1 (1,5)
Protokollverletzung	1 (0,8)	1 (1,5)
Entzug der Einverständniserklärung	5 (3,9)	2 (3,1)
Andere Gründe	1 (0,8)	0
Abbruch der Einnahme der Studienmedikation, n (%)	10 (7,8)	8 (12,3)
Aufgrund von:		
UE	4 (3,1)	4 (6,2)
Klinisches PBC-Ereignis	0 <sup>2)</sup>	0
Lost to Follow-up	1 (0,8)	1 (1,5)
Protokollverletzung	1 (0,8)	1 (1,5)
Entzug der Einverständniserklärung	4 (3,1)	2 (3,1)
Teilnahme an ASSURE zugestimmt, n (%)	110 (85,9)	54 (83,1)
RESPONSE abgeschlossen und für ASSURE geeignet <sup>3)</sup>	106 (82,8) <sup>4)</sup>	54 (83,1) <sup>4)</sup>
Teilnahme an ASSURE abgelehnt, n (%)	18 (14,1)	11 (16,9)
RESPONSE inkl. Sicherheits-FU abgeschlossen	7 (5,5)	3 (4,6)
RESPONSE ohne Sicherheits-FU abgeschlossen	1 (0,8)	0
Durchschnittliche tägliche Dosis <sup>5)</sup>		
MW (SD)	9,81 (0,68)	9,9 (0,71)
Median (min; max)	9,95 (5,8; 12,5)	9,92 (8; 12,7)
Studiendosis, n (%) <sup>6)</sup>		
10 mg	127 (99,2)	65 (100)
5 mg	1 (0,8)	0
Dosisunterbrechungen, n (%)	12 (9,4)	5 (7,7)
1	11 (8,6)	5 (7,7)
2	1 (0,8)	0
Mediane Behandlungsdauer in Wochen (min; max)	52,14 (5,4; 54,7)	52 (1,3; 55,4)
Behandlung mit UDCA, n (%)	120 (93,8)	62 (95,4)
Mediane Behandlungsdauer in Wochen (min; max)	52,14 (5; 59)	52,14 (6,3; 55,4)

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

- <sup>2)</sup> Eine Person hatte ein klinisches PBC-Ereignis (Hospitalisierung) 7 Tage nach Abschluss der Behandlungsphase. Gemäß Präspezifizierung sollte die Person die Studie abbrechen, dies geht jedoch nicht eindeutig aus den Studienunterlagen hervor.
- <sup>3)</sup> Russische Teilnehmende (n = 8 im Seladelpar-Arm und n = 1 im Placebo-Arm) waren aus logistischen Gründen nicht für die Teilnahme an der ASSURE-Studie geeignet.
- <sup>4)</sup> Prozentangaben aus eigener Berechnung. Gemäß EPAR gingen insgesamt 159 statt 160 Personen in die LTE über.
- <sup>5)</sup> Durchschnittliche tägliche Dosis = Kumulative Dosis / Gesamttag der Exposition; Kumulative Dosis = [(Anzahl der ausgegebenen Kapseln – die größere Zahl von (zurückgegebenen Kapseln, Anzahl der verpassten Dosen)) × verabreichte Dosis pro Kapsel] für jede Ausgabe des Medikaments bzw. jeden Besuch. Die Definition der „Gesamttag der Exposition“ ist nicht bekannt. Wenn keine Informationen zu zurückgegebenen oder versäumten Dosen verfügbar waren, wurde angenommen, dass alle geplanten Dosen eingenommen wurden. 2 Personen wurden laut CSR aus der Berechnung ausgeschlossen, da aufgrund einer großen Anzahl nicht zurückgegebener Dosen keine genaue Schätzung der Dosierung durchgeführt werden konnte.
- <sup>6)</sup> 1 Person (0,8 %) im Seladelpar-Arm wurde im Verlauf der Studie auf 5 mg heruntertitriert, nachdem es zu einer Unterbrechung der Studienmedikation aufgrund eines TEAE Grad 2 (medikamenteninduzierte Leberschädigung) kam. Ein Zusammenhang des Ereignisses mit der Studienmedikation wurde als unwahrscheinlich bewertet, und die Dosis etwa 2 Monate nach Abklingen des Ereignisses wieder auf 10 mg erhöht.

Abkürzungen: CSR: Clinical Study Report; EPAR: European Public Assessment Report; FU: Follow-Up; ITT: Intention to Treat; MW: Mittelwert; LTE: Langzeitnachsichtungsstudie; PBC: Primär biliäre Cholangitis; SD: Standardabweichung; TEAE: Treatment Emergent Adverse Event; UDCA: Ursodeoxycholsäure; UE: Unerwünschtes Ereignis.

*Tabelle 10: Charakterisierung der Studienpopulation; Studie RESPONSE, ITT-Population*

<b>Studie RESPONSE Charakterisierung der Studienpopulation</b>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>
<i>Alter (Jahre)</i> MW (SD) Median (min; max)	56,6 (9,99) 57,0 (28,0; 75,0)	57,0 (9,17) 58,0 (33,0; 75,0)
<i>Altersgruppe (Jahre), n (%)</i> < 65 Jahre ≥ 65 Jahre	99 (77,3) 29 (22,7)	53 (81,5) 12 (18,5)
<i>Geschlecht, n (%)</i> weiblich männlich	123 (96,1) 5 (3,9)	60 (92,3) 5 (7,7)
<i>Abstammung (genetisch), n (%)</i> Amerikanische oder Alaska-Ureinwohner Asiatisch Schwarz oder Afroamerikanisch Weiß/Kaukasisch Nicht angegeben	3 (2,3) 7 (5,5) 2 (1,6) 114 (89,1) 2 (1,6)	3 (4,6) 4 (6,2) 2 (3,1) 56 (86,2) 0
<i>Region, n (%)</i> Asien-Pazifik (APAC) Europa, Naher Osten und Afrika (EMEA) Europa Naher Osten Afrika Lateinamerika Nordamerika	5 (3,9) 49 (38,3) k. A. k. A. k. A. 24 (18,8) 50 (39,1)	6 (9,2) 27 (41,5) k. A. k. A. k. A. 19 (29,2) 13 (20,0)
<i>Ethnische Zugehörigkeit, n (%)</i> Hispanisch oder lateinamerikanisch Nicht-hispanisch oder lateinamerikanisch fehlend	29 (22,7) 97 (75,8) 2 (1,6)	27 (41,5) 38 (58,5) 0

<b>Studie RESPONSE Charakterisierung der Studienpopulation</b>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>
<i>BMI (kg/m<sup>2</sup>)</i> MW (SD) Median (min; max)	27,2 (5,61) 26,0 (17,5; 45,0)	26,8 (4,81) 26,2 (17,4; 40,1)
<i>eGFR (mL/min/1,73 m<sup>2</sup>)</i> MW (SD) Median (min; max) Q1, Q3	98,9 (23,7) 95,8 (44,2; 189,1) 81,8, 113,2	98,5 (22,3) 96,6 (55,9; 160,2) 81,9, 112,2
<b>Medikamentöse Behandlung zu Studienbeginn und Vorbehandlungen</b>		
<i>Behandlung mit UDCA zu Baseline, n (%)</i> Ja Nein <sup>1)</sup>	120 (93,8) 8 (6,3)	61 (93,8) 4 (6,2)
<i>Dauer der UDCA-Behandlung bis Baseline (Jahre)</i> MW (SD) Median (min; max)	7,3 (6,46) 4,7 (0,0; 33,0)	7,8 (6,24) 6,0 (0; 28,0)
<i>Vorbehandlung mit OCA und/oder Fibraten, n (%)</i> Ja Nein	20 (15,6) 108 (84,4)	13 (20,0) 52 (80,0)
<b>Krankheitsmerkmale</b>		
<i>AMA-Status, n (%)</i> Positiv <sup>2)</sup> Negativ Mehrdeutig	106 (82,8) 20 (15,6) 2 (1,6)	55 (84,6) 9 (13,8) 1 (1,5)
<i>Pruritus NRS</i> MW (SD) Median (min; max) Q1; Q3	3,0 (2,81) 2,8 (0; 9) 0,1; 5,3	3,0 (2,96) 1,8 (0; 9) 0,2; 5,6
<i>Pruritus-NRS-Level, n (%)<sup>3)</sup></i> < 4 ≥ 4	79 (61,7) 49 (38,3)	42 (64,6) 23 (35,4)
<i>Fatigue, n (%)</i> Ja Nein	60 (46,9) 68 (53,1)	33 (50,8) 32 (49,2)
<i>Zirrhosis zu Baseline<sup>4)</sup>, n (%)</i> Ja <sup>5)</sup> Nein	18 (14,1) 110 (85,9)	9 (13,8) 56 (86,2)
<i>Child-Pugh-Score</i> MW (SD) Median (min; max)	5,1 (0,32) 5,0 (5,0; 6,0)	5,0 (0) 5,0 (5,0; 5,0)
<i>Portale Hypertension, n (%)</i> Ja Nein	0 128 (100,0)	3 (4,6) 62 (95,4)
<i>Rotterdam Krankheitsstadium<sup>6)</sup>, n (%)</i> Mild Mäßig fortgeschritten	106 (82,8) 22 (17,2)	60 (92,3) 5 (7,7)

<b>Studie RESPONSE</b> <b>Charakterisierung der Studienpopulation</b>	<b>Seladelpar</b> N = 128	<b>Placebo</b> N = 65
<i>MELD-Score</i> MW (SD) Median (min; max)	6,8 (0,97) 6,3 (6,0; 11,7)	6,7 (0,93) 6,5 (6,0; 11,0)
<i>Lebersteifigkeit (kPa), n (%)</i> Median (min; max) <i>Fibrose-Score gemäß Lebersteifigkeit, n (%)</i>	115 8,0 (3,1; 43,2)	62 7,5 (3,8; 23,0)
F0 Keine Fibrose (< 7,1 kPa)	44 (38,3)	26 (41,9)
F1 Geringgradige Fibrose (7,1 – < 8,8 kPa)	22 (19,1)	15 (24,2)
F2 Mittelgradige Fibrose (8,8 – < 10,7 kPa)	17 (14,8)	7 (11,3)
F3 Hochgradige Fibrose (10,7 – < 16,9 kPa)	21 (18,3)	10 (16,1)
F4 Zirrhose (≥ 16,9 kPa)	11 (9,6)	4 (6,5)
<i>Enhanced-Liver-Fibrosis(ELF)-Score</i> MW (SD) Median (min; max)	10,2 (1,03) 10,1 (8,1; 13,3)	10,2 (0,85) 10,0 (8,6; 12,3)
<i>ALP (U/L)<sup>3)</sup></i> MW (SD) Median (min; max)	314,6 (122,96) 278,3 (182,3; 786,3)	313,8 (117,68) 281,3 (160,7; 697,7)
<i>Gesamtbilirubin</i> MW (SD) Median (min; max)	0,8 (0,31) 0,7 (0,3; 1,9)	0,7 (0,31) 0,7 (0,3; 1,9)
<i>ALT (U/L)</i> MW (SD) Median (min; max)	47,4 (23,47) 45,3 (13,0; 108,8)	48,2 (22,83) 43,0 (9,3; 115,3)
<i>AST (U/L)</i> MW (SD) Median (min; max)	39,6 (16,14) 37,7 (15,7; 94,0)	41,7 (16,03) 38,0 (16,0; 84,0)
<i>GGT (U/L)</i> MW (SD) Median (min; max)	269,0 (240,04) 195,2 (12,7; 1402,0)	287,5 (249,56) 216,3 (41,5; 1088,0)
<i>Thrombozytenzahl (10<sup>3</sup>/μL)</i> MW (SD) Median (min; max) Q1; Q3	125 241,7 (78,87) 241,3 (90,0; 477,0) 178,3; 287,0	65 241,9 (84,46) 232,0 (102,7; 506,0) 184,7; 309,7
<i>INR</i> MW (SD) Median (min; max)	1,0 (0,08) 1,0 (0,8; 1,3)	1,0 (0,09) 1,0 (0,9; 1,5)
<i>Albumin (g/dL)</i> MW (SD) Median (min; max)	4,2 (0,27) 4,2 (3,0; 4,8)	4,1 (0,23) 4,1 (3,6; 4,6)

<sup>1)</sup> Alle Personen ohne UDCA-Behandlung zu Baseline waren UDCA-intolerant.

<sup>2)</sup> Positiv > 1:40 in der Immunfluoreszenz oder M2 positiv im ELISA.

<sup>3)</sup> Stratifizierungsfaktor für die Randomisierung.

<sup>4)</sup> Diagnosekriterien (mindestens eines musste erfüllt sein):

- frühere Leberbiopsie, die eine Zirrhose anzeigt;
- akute oder frühere dekompensierte Lebererkrankung oder andere klinischer Zustände, die auf eine Leberzirrhose und/oder portale Hypertension hinweisen;
- Lebersteifigkeit > 16,9 kPa zum Screening;

- Kombination aus Thrombozytenzahl  $< 140 \times 10^3 / \mu\text{l}$  mit Serumalbumin  $< 3,5 \text{ g/dl}$  und/oder INR  $> 1,3$  und/oder Gesamtbilirubin  $> 1,0 \times \text{ULN}$ ;
- radiologische Hinweise auf eine Leberzirrhose mit gleichzeitiger Splenomegalie;
- klinische Beurteilung durch die Prüffärztin/den Prüffarzt.

<sup>5)</sup> Alle Child-Pugh Class „A“.

<sup>6)</sup> Mild: normales Gesamtbilirubin und normales Albumin; mäßig fortgeschritten: entweder abnormes Albumin oder abnormes Gesamtbilirubin; fortgeschritten: sowohl abnormes Albumin als auch abnormes Gesamtbilirubin.

Abkürzungen: AMA: antimitochondriale Antikörper; ALP: Alkalische Phosphatase; ALT: Alanin-Aminotransferase; APAC: Asien-Pazifik; AST: Aspartat-Aminotransferase; BMI: Body-Mass-Index; eGFR: estimated Glomerular Filtration Rate; ELISA: Enzyme-Linked Immunosorbent Assay; ELF: Fortgeschrittene Leberfibrose; GGT: Gamma-Glutamyltransferase; INR: International Normalized Ratio; ITT: Intention to Treat; MW: Mittelwert; kPa: Kilopascal; MELD: Model for End-Stage Liver Disease; MW: Mittelwert; NRS: Numerical Rating Scale; OCA: Obeticholsäure; SD: Standardabweichung; UDCA: Ursodeoxycholsäure; ULN: Oberer Normwert.

### Protokollverletzungen

In beiden Studienarmen trat bei etwa der Hälfte der Teilnehmenden eine signifikante Protokollverletzung auf. Die meisten Abweichungen betrafen die Verabreichung/Ausgabe der Studienmedikation (20,7 %), die Studienverfahren/-bewertungen (17,6 %), den Prozess/Zeitpunkt der Einwilligungserklärung (13 %), die Terminierung der Visiten (13 %) sowie die Therapieadhärenz (10,9 %). Ein höherer Anteil der Teilnehmenden im Placebo-Arm zeigte Abweichungen bei der Terminierung der Visiten im Vergleich zum Seldelpar-Arm (18,5 % vs. 10,2 %). Umgekehrt traten in der Seladelpar-Gruppe häufiger Abweichungen bei der Randomisierung der Studienbehandlung auf als in der Placebo-Gruppe (8,6 % vs. 1,5 %); siehe auch Kapitel 2.5.

Anhand der in der Studie RESPONSE dokumentierten Protokollverletzungen ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

### Begleitmedikation

Insgesamt erhielten 97,7 % der Teilnehmenden im Seladelpar-Arm und 98,5 % im Placebo-Arm im Verlauf der Studie mindestens eine Begleitmedikation. Die Verteilung der Begleitmedikationen war in beiden Gruppen weitgehend vergleichbar, mit Ausnahme der Lipid-modifizierenden Substanzen: Diese wurden im Seladelpar-Arm von 32,8 % und im Placebo-Arm von 52,3 % der Teilnehmenden eingenommen, was einem Unterschied von etwa 20 % entspricht. Vorliegend wird nicht davon ausgegangen, dass diese Imbalance relevante Auswirkungen auf die Ergebnisse hat. Systemische Kortikosteroide wurden bei 10,2 % in der Seladelpar-Gruppe und bei 16,9 % in der Placebo-Gruppe eingesetzt; es bleibt unklar, ob diese Anwendung jeweils länger als 2 Wochen erfolgte, was laut Protokoll nicht zulässig gewesen wäre. Rund 7 % der Teilnehmenden erhielt begleitend Cholestyramin, das gemäß European Public Assessment Report (EPAR) zur Behandlung von PBC-assoziiertem Pruritus eingesetzt wird; dabei bestand ein leichter Unterschied zwischen den Gruppen (4,7 % in der Seladelpar- und 12 % in der Placebo-Gruppe). Rifampicin, das laut EPAR ebenfalls (off-label) bei cholestatischem Pruritus angewandt wird, wurde in beiden Studienarmen bei knapp 3 % der Teilnehmenden eingesetzt. Etwa 94 % der Teilnehmenden in beiden Studienarmen wurden während der Studie mit UDCA behandelt (siehe Tabelle 10).

## 3.2 Mortalität

In der Studie RESPONSE sind bis zum Ende der Studie vom 11.08.2023 keine Todesfälle aufgetreten. Die mediane Beobachtungsdauer für die Sicherheitsüberwachung lag im Seladelpar-Arm bei 52,4 Wochen (min: 4,4; max: 58,3) und im Placebo-Arm bei 52,1 Wochen (min: 5,6; max: 58,1).

## 3.3 Morbidität

### Pruritus

#### Pruritus NRS

Die Pruritus NRS wurde in Form eines e-Tagebuchs erhoben (siehe Kapitel 2.3.2). Daten zur Vollständigkeit der Tagebücher liegen nicht vor.

Die mediane Beobachtungsdauer für die Pruritus NRS lag in beiden Studienarmen bei 46,6 Wochen (jeweils min: 0,1; max: 46,6). Bei der Interpretation der Ergebnisse zu Monat 12 (siehe Tabelle 12 und Tabelle 14) ist somit zu berücksichtigen, dass keine tatsächlichen Werte zu Woche 52 vorlagen.

*Tabelle 11: Veränderung der Pruritus NRS zu Monat 6 gegenüber Baseline auf Basis der gemittelten 7-Tage-Intervalle, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>1)</sup>*

Studie RESPONSE Pruritus NRS <sup>2)</sup> zu Monat 6	Seladelpar N = 128	Placebo N = 65
<i>Baseline</i> <sup>3)</sup>		
n (%)	128 (100)	65 (100)
MW (SD)	3,03 (2,81)	3,02 (2,96)
<i>Monat 6</i>		
n (%)	121 (94,5)	61 (93,8)
MW (SD)	1,62 (1,88)	2,44 (2,35)
<i>Veränderung von Baseline zu Monat 6</i>		
n (%) <sup>4)5)</sup>	127 (99,2) <sup>7)</sup>	64 (98,5) <sup>7)</sup>
LS Mean (SE) <sup>6)</sup>	-1,33 (0,14)	-0,42 (0,19)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup>	-0,90 [-1,35; -0,45]; 0,0001	
SMD [95%-KI] <sup>8)</sup>	-0,58 [-0,89; -0,28]	

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> Intensität des schlimmsten Juckreizes innerhalb der letzten 24 Stunden auf einer 11-stufigen Skala von 0 (= kein Juckreiz) bis 10 (= schlimmster vorstellbarer Juckreiz) (Selbstbericht).

<sup>3)</sup> Baseline war definiert als arithmetisches Mittel aller täglich dokumentierten Werte während der Run-in-Phase (ca. 2 Wochen vor Tag 1) und an Tag 1.

<sup>4)</sup> Der Umgang mit fehlenden Werten ist Kapitel 2.4 zu entnehmen.

<sup>5)</sup> Anzahl an Personen, die sowohl einen Baseline-Wert als auch mindestens einen Wert nach Baseline aufwiesen und somit in das MMRM eingeschlossen wurden.

<sup>6)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-NRS, Randomisierungsstratum (ALP-Wert < 350 U/L versus ALP-Wert ≥ 350 U/L), Behandlungsgruppe, Woche sowie die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline-Wert wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als zusätzlicher Faktor in das Modell aufgenommen.

<sup>7)</sup> Prozentangaben aus eigener Berechnung.

<sup>8)</sup> Hedges g, post-hoc berechnet.

Abkürzungen: ALP: Alkalische Phosphatase; ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MMRM: Gemischtes Modell mit Messwiederholungen; MW: Mittelwert; NRS: Numerical Rating Scale; SD: Standardabweichung; SE: Standardfehler.

**Tabelle 12: Veränderung der Pruritus NRS zu Monat 12<sup>1)</sup> gegenüber Baseline auf Basis der gemittelten 7-Tage-Intervalle (ergänzend dargestellt), MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>2)</sup>**

Studie RESPONSE Pruritus NRS <sup>3)</sup> zu Monat 12 (ergänzend dargestellt)	Seladelpar N = 128	Placebo N = 65
<i>Baseline</i> <sup>4)</sup> n (%) MW (SD)	128 (100) 3,03 (2,81)	65 (100) 3,02 (2,96)
<i>Monat 12</i> n (%) MW (SD)	105 (82) 1,55 (1,94)	48 (73,8) 2,21 (2,43)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>5)6)</sup> LS Mean (SE) <sup>7)</sup>	127 (99,2) <sup>8)</sup> -1,35 (0,16)	64 (98,5) <sup>8)</sup> -0,52 (0,22)
LS Mean Difference [95%-KI]; p-Wert <sup>7)</sup> SMD [95%-KI] <sup>9)</sup>		-0,83 [-1,35; -0,31]; 0,002 -0,47 [-0,78; -0,17]

<sup>1)</sup> Die mediane Beobachtungsdauer für die Pruritus NRS lag in beiden Studienarmen bei 46,6 Wochen (jeweils min: 0,1; max: 46,6).

<sup>2)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>3)</sup> Intensität des schlimmsten Juckreizes auf einer 11-stufigen Skala von 0 (= kein Juckreiz) bis 10 (= schlimmster vorstellbarer Juckreiz) innerhalb der letzten 24 Stunden (Selbstbericht).

<sup>4)</sup> Baseline war definiert als arithmetisches Mittel aller täglich dokumentierten Werte während der Run-in-Phase (ca. 2 Wochen vor Tag 1) und an Tag 1.

<sup>5)</sup> Der Umgang mit fehlenden Werten ist Kapitel 2.4 zu entnehmen.

<sup>6)</sup> Anzahl an Personen, die sowohl einen Baseline-Wert als auch mindestens einen Wert nach Baseline aufwiesen und somit in das MMRM eingeschlossen wurden.

<sup>7)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-NRS, Randomisierungsstratum zu Studienbeginn (ALP-Wert < 350 U/L versus ALP-Wert ≥ 350 U/L), Behandlungsgruppe, Woche sowie die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline-NRS wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als zusätzlicher Faktor in das Modell aufgenommen.

<sup>8)</sup> Prozentangaben aus eigener Berechnung.

<sup>9)</sup> Hedges g, post-hoc berechnet.

Abkürzungen: ALP: Alkalische Phosphatase; ITT: Intention to Treat; KI: Konfidenzintervall; LS: Least Squares; MMRM: Gemischtes Modell mit Messwiederholungen; MW: Mittelwert; NRS: Numerical Rating Scale; SD: Standardabweichung; SE: Standardfehler; SMD: Standardisierte Mittelwertdifferenz.

Für die MMRM-Analyse zu Monat 6 zeigen die in Modul 4 dargestellten Subgruppenanalysen keine signifikanten Interaktionsterme.

**Tabelle 13: Pruritus-NRS-Ansprechen zu Monat 6 gegenüber Baseline<sup>1)</sup> bei Teilnehmenden mit Pruritus NRS > 0 zu Baseline (auf Basis der gemittelten 7-Tage-Intervalle), Responderanalysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>2)</sup>**

<b>Studie RESPONSE Pruritus NRS<sup>3)</sup> – Responder zu Monat 6</b>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>	<b>RR [95%-KI]; p-Wert</b>
Personen mit Baseline > 0 <sup>4)</sup> , n (%)	98 (76,6) <sup>7)</sup>	50 (76,9) <sup>7)</sup>	
Verbesserung um $\geq 4$ oder auf 0 Ereignisse <sup>5)</sup>	28 (21,9) <sup>7)</sup>	10 (15,4) <sup>7)</sup>	k. A.
Zensierung <sup>6)</sup>	k. A.	k. A.	

<sup>1)</sup> Baseline war definiert als arithmetisches Mittel aller täglich dokumentierten Werte während der Run-in-Phase (ca. 2 Wochen vor Tag 1) und an Tag 1.

<sup>2)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>3)</sup> Intensität des schlimmsten Juckreizes innerhalb der letzten 24 Stunden auf einer 11-stufigen Skala von 0 (= kein Juckreiz) bis 10 (= schlimmster vorstellbarer Juckreiz) (Selbstbericht).

<sup>4)</sup> Im vorliegenden Fall wurden die Prozentangaben (Eigenberechnung) auf die für die Nutzenbewertung relevante ITT-Population bezogen. Diese umfasst auch Personen mit einem Pruritus-NRS-Wert von 0 zu Baseline, welche in einer Responderanalyse als Non-Responder gewertet werden sollten.

<sup>5)</sup> Personen, für die keine Bewertung vorliegt, die die Behandlung vor Monat 6 abgebrochen haben oder aus anderen Gründen fehlende Werte aufweisen, werden als Non-Responder gewertet.

<sup>6)</sup> Den Studienunterlagen konnte nicht entnommen werden, wie viele der Personen ohne Ereignis tatsächlich Non-Responder waren und wie viele zu Monat 6 fehlende Werte aufwiesen oder die Studie vorzeitig abgebrochen haben. Über den gesamten Studienzeitraum hinweg brachen 11 Personen im Seladelpar-Arm und 8 Personen im Placebo-Arm die Studie vorzeitig ab. Allerdings ist unklar, wie viele dieser Abbrüche bereits innerhalb der ersten 6 Monate stattgefunden haben.

<sup>7)</sup> Prozentangaben aus eigener Berechnung bezogen auf die ITT.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; NRS: Numerical Rating Scale; RR: Relatives Risiko.

**Tabelle 14: Pruritus-NRS-Ansprechen zu Monat 12<sup>1)</sup> gegenüber Baseline<sup>2)</sup> bei Teilnehmenden mit Pruritus NRS > 0 zu Baseline (auf Basis der gemittelten 7-Tage-Intervalle) (ergänzend dargestellt), Responderanalysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>3)</sup>**

<b>Studie RESPONSE Pruritus NRS<sup>4)</sup> – Responder zu Monat 12 (ergänzend dargestellt)</b>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>	<b>RR [95%-KI]; p-Wert</b>
Personen mit Baseline > 0 <sup>5)</sup> , n (%)	98 (76,6) <sup>8)</sup>	50 (76,9) <sup>8)</sup>	
Verbesserung um $\geq 4$ oder auf 0 Ereignisse <sup>6)</sup>	24 (18,8) <sup>8)</sup>	9 (13,8) <sup>8)</sup>	k. A.
Zensierung <sup>7)</sup>	k. A.	k. A.	

<sup>1)</sup> Die mediane Beobachtungsdauer für die Pruritus NRS lag in beiden Studienarmen bei 46,6 Wochen (jeweils min: 0,1; max: 46,6).

<sup>2)</sup> Baseline war definiert als arithmetisches Mittel aller täglich dokumentierten Werte während der Run-in-Phase (ca. 2 Wochen vor Tag 1) und an Tag 1.

<sup>3)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>4)</sup> Intensität des schlimmsten Juckreizes auf einer 11-stufigen Skala von 0 (= kein Juckreiz) bis 10 (= schlimmster vorstellbarer Juckreiz) innerhalb der letzten 24 Stunden (Selbstbericht).

<sup>5)</sup> Im vorliegenden Fall wurden die Prozentangaben (Eigenberechnung) auf die für die Nutzenbewertung relevante ITT-Population bezogen. Diese umfasst auch Personen mit einem Pruritus-NRS-Wert von 0 zu Baseline, welche in einer Responderanalyse als Non-Responder gewertet werden sollten.

- <sup>6)</sup> Personen, für die keine Bewertung vorliegt, die die Behandlung vor Monat 12 abgebrochen haben oder aus anderen Gründen fehlende Werte aufweisen, werden als Non-Responder gewertet.
- <sup>7)</sup> Insgesamt brachen im Seladelpar-Arm 11 und im Placebo-Arm 8 Personen die Studie vorzeitig ab und wurden zu Monat 12 somit als Non-Responder gezählt. Den Studienunterlagen konnte nicht entnommen werden, wie viele der restlichen Personen ohne Ereignis tatsächlich Non-Responder waren und wie viele zu Monat 12 fehlende Werte aufwiesen.
- <sup>8)</sup> Prozentangaben aus eigener Berechnung bezogen auf die ITT.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; NRS: Numerical Rating Scale; RR: Relatives Risiko.

Für die Responderanalyse zu Monat 6 zeigen die in Modul 4 dargestellten Subgruppenanalysen keine signifikanten Interaktionsterme.

### 5-D-Itch-Scale

Analog zur Pruritus NRS wurde die 5-D Itch anhand eines e-Tagebuchs erhoben. Daten zur Vollständigkeit des Tagebuchs liegen nicht vor. Aufgrund der in Kapitel 2.3.2 beschriebenen Unsicherheiten, werden zusätzlich zum Gesamtscore die Domänen dargestellt.

Die mediane Beobachtungsdauer für die 5-D Itch lag in beiden Studienarmen bei 44 Wochen (für Seladelpar min: 4,1; max: 46; für Placebo min: 0,1; max: 45,1). Bei der Interpretation der Ergebnisse zu Monat 12 ist somit zu berücksichtigen, dass keine tatsächlichen Werte zu Woche 52 vorlagen.

Tabelle 15: Veränderung der 5-D Itch zu Monat 12 gegenüber Baseline, MMRM-Analysen<sup>1)</sup>; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>2)</sup>

Studie RESPONSE 5-D Itch <sup>3)</sup>	Seladelpar N = 128	Placebo N = 65
<b>Gesamtscore</b>		
<i>Baseline</i> <sup>4)</sup>		
n (%)	128 (100)	65 (100)
MW (SD)	11,6 (4,85)	11,2 (4,65)
<i>Monat 12</i>		
n (%)	96 (75)	46 (70,8)
MW (SD)	8,83 (3,68)	10,7 (4,38)
<i>Veränderung von Baseline zu Monat 12</i>		
n (%) <sup>5)</sup>	128 (100) <sup>7)</sup>	64 (98,5) <sup>7)</sup>
LS Mean (SE) <sup>6)</sup>	-2,26 (0,33)	0,05 (0,46)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup>	-2,31 [-3,40; -1,23]; < 0,0001	
SMD [95%-KI] <sup>8)</sup>	-0,62 [-0,93; -0,31]	
<b>Domäne „Grad“</b>		
<i>Baseline</i> <sup>4)</sup>		
n (%)	128 (100)	65 (100)
MW (SD)	2,26 (1,04)	2,24 (0,99)
<i>Monat 12</i>		
n (%)	96 (75)	46 (70,8)
MW (SD)	1,71 (0,81)	2,04 (0,84)
<i>Veränderung von Baseline zu Monat 12</i>		
n (%) <sup>5)</sup>	128 (100) <sup>7)</sup>	64 (98,5) <sup>7)</sup>
LS Mean (SE) <sup>6)</sup>	-0,42 (0,07)	-0,05 (0,10)

<b>Studie RESPONSE 5-D Itch<sup>3)</sup></b>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup> SMD [95%-KI] <sup>8)</sup>	-0,38 [-0,61; -0,15]; 0,002 -0,47 [-0,77; -0,16]	
<b>Domäne „Dauer“</b>		
<i>Baseline<sup>4)</sup></i> n (%) MW (SD)	128 (100) 1,59 (1,11)	65 (100) 1,56 (1,08)
<i>Monat 12</i> n (%) MW (SD)	96 (75) 1,16 (0,59)	46 (70,8) 1,54 (0,98)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>5)</sup> LS Mean (SE) <sup>6)</sup>	128 (100) <sup>7)</sup> -0,32 (0,07)	64 (98,5) <sup>7)</sup> 0,08 (0,10)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup> SMD [95%-KI] <sup>8)</sup>	-0,40 [-0,63; -0,16]; 0,001 -0,49 [-0,79; -0,19]	
<b>Domäne „Beeinträchtigung“</b>		
<i>Baseline<sup>4)</sup></i> n (%) MW (SD)	128 (100) 2,27 (1,37)	65 (100) 2,22 (1,24)
<i>Monat 12</i> n (%) MW (SD)	96 (75) 1,57 (0,96)	46 (70,8) 2,09 (1,35)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>5)</sup> LS Mean (SE) <sup>6)</sup>	128 (100) <sup>7)</sup> -0,55 (0,09)	64 (98,5) <sup>7)</sup> 0,05 (0,13)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup> SMD [95%-KI] <sup>8)</sup>	-0,60 [-0,91; -0,29]; 0,0002 -0,57 [-0,87; -0,26]	
<b>Domäne „Verlauf“</b>		
<i>Baseline<sup>4)</sup></i> n (%) MW (SD)	128 (100) 3,35 (1,09)	65 (100) 3,13 (1,16)
<i>Monat 12</i> n (%) MW (SD)	96 (75) 2,64 (1,35)	46 (70,8) 3,04 (1,25)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>5)</sup> LS Mean (SE) <sup>6)</sup>	128 (100) <sup>7)</sup> -0,64 (0,13)	64 (98,5) <sup>7)</sup> -0,17 (0,18)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup> SMD [95%-KI] <sup>8)</sup>	-0,47 [-0,90; -0,04]; 0,03 -0,32 [-0,62; -0,02]	
<b>Domäne „Lokalisation“</b>		
<i>Baseline<sup>4)</sup></i> n (%) MW (SD)	128 (100) 2,16 (1,29)	65 (100) 2,08 (1,20)

Studie <b>RESPONSE 5-D Itch</b> <sup>3)</sup>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>
<i>Monat 12</i> n (%) MW (SD)	96 (75) 1,76 (1,19)	46 (70,8) 1,98 (1,22)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>5)</sup> LS Mean (SE) <sup>6)</sup>	128 (100) <sup>7)</sup> -0,25 (0,09)	64 (98,5) <sup>7)</sup> 0,14 (0,12)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup> SMD [95%-KI] <sup>8)</sup>	-0,38 [-0,67; -0,10]; 0,009 -0,39 [-0,69; -0,09]	

<sup>1)</sup> Die kontinuierlichen Analysen mittels MMRM werden als post-hoc eingestuft.

<sup>2)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>3)</sup> Die zu erreichende Gesamtpunktzahl liegt zwischen 5 und 25 Punkten. Jede Domäne kann einen Wert von 1 bis 5 Punkten erreichen. Höhere Punkte deuten auf einen stärkeren Juckreiz hin.

<sup>4)</sup> Baseline war definiert als arithmetisches Mittel aller verfügbaren Messungen bei Screening, während der Run-in-Phase (ca. 2 Wochen vor Tag 1), an Tag 1 sowie bei ungeplanten Visiten vor oder an Tag 1.

<sup>5)</sup> Anzahl an Personen, die sowohl einen Baseline-Wert als auch mindestens einen Wert nach Baseline aufwiesen und somit in das MMRM eingeschlossen wurden.

<sup>6)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-Werten, Randomisierungsstratum zu Studienbeginn (ALP-Wert < 350 U/L vs. ALP-Wert ≥ 350 U/L und Pruritus NRS < 4 vs. Pruritus NRS ≥ 4), Behandlungsgruppe, Woche und die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline-Wert wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als zusätzlicher Faktor in das Modell aufgenommen.

<sup>7)</sup> Prozentangaben aus eigener Berechnung.

<sup>8)</sup> Hedges g, post-hoc berechnet.

Abkürzungen: ALP: Alkalische Phosphatase; ITT: Intention to Treat; KI: Konfidenzintervall; LS: Least Squares; MMRM: Gemischtes Modell mit Messwiederholungen; MW: Mittelwert; NRS: Numerical Rating Scale; SD: Standardabweichung; SMD: Standardisierte Mittelwertdifferenz.

Für die dargestellte MMRM-Analyse des Gesamtscores zu Monat 12 zeigten die durchgeführten Subgruppenanalysen in Modul 4 eine signifikante Interaktion für das Gesamtbilirubin (p = 0,02). Bei Personen mit einem Gesamtbilirubin ≥ 0,6 x ULN zu Baseline ergab sich ein statistisch signifikanter Vorteil für Seladelpar (LS Mean Difference [95%-KI]: -3,56 [-4,97; -2,15]; p < 0,0001). Bei Personen mit einem Gesamt-Bilirubin < 0,6 x ULN zu Baseline war der Effekt gleichgerichtet, jedoch nicht signifikant (LS Mean Difference [95%-KI]: -1,02 [-2,71; 0,66], p = 0,2).

In der Domäne „Grad“ gab es ebenfalls eine signifikante Effektmodifikation für das Merkmal „Gesamtbilirubin (< 0,6 x ULN, ≥ 0,6 x ULN)“ (p = 0,04), in der Domäne „Dauer“ für das Merkmal „Alter zum Screening“ (p = 0,04) und „Pruritus NRS“ (p = 0,02), in der Domäne „Verlauf“ für das Merkmal „Vorherige Einnahme von OCA und/oder Fibraten“ (p = 0,04) sowie in der Domäne „Lokalisation“ für das Merkmal „Pruritus NRS“ (p = 0,003). In der Domäne „Beeinträchtigung“ zeigten sich keine signifikanten Effektmodifikationen.

In der Gesamtschau zeigen sich keine konsistenten, signifikanten Interaktionsterme für dasselbe Subgruppenmerkmal über die verschiedenen Domänen der 5-D Itch hinweg.

### MELD-Score $\geq 15$ bei mindestens 2 aufeinanderfolgenden Visiten

Bis zum Ende der Studie am 11.08.2023 trat laut prüfärztlicher Einschätzung sowie gemäß CERC-Bewertung in keinem der beiden Studienarme ein MELD-Score  $\geq 15$  bei mindestens 2 aufeinanderfolgenden Visiten auf. Die mediane Beobachtungsdauer für den MELD-Score ist nicht bekannt.

## 3.4 Lebensqualität

### PBC-40

Die mediane Beobachtungsdauer für den PBC-40 lag im Seladelpar-Arm bei 51,9 Wochen (min: 0,1; max: 54,3) und im Placebo-Arm bei 52 Wochen (min: 0,1; max: 55,4).

Tabelle 16: Veränderung des PBC-40 zu Monat 12 gegenüber Baseline, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>1)</sup>

Studie RESPONSE PBC-40 <sup>2)</sup>	Seladelpar N = 128	Placebo N = 65
<b>Domäne „Juckreiz“</b>		
<i>Baseline</i> <sup>3)</sup>		
n (%)	128 (100)	65 (100)
MW (SD)	5,14 (3,85)	5,60 (3,97)
<i>Monat 12</i>		
n (%)	94 (73,4)	51 (78,5)
MW (SD)	3,51 (3,33)	4,31 (3,40)
<i>Veränderung von Baseline zu Monat 12</i>		
n (%) <sup>4)</sup>	126 (98,4) <sup>6)</sup>	63 (96,9) <sup>6)</sup>
LS Mean (SE) <sup>5)</sup>	-1,31 (0,26)	-0,48 (0,36)
LS Mean Difference [95%-KI]; p-Wert <sup>5)</sup>	-0,83 [-1,68; 0,02]; 0,054	
<b>Domäne „Generelle Symptome“</b>		
<i>Baseline</i> <sup>3)</sup>		
n (%)	128 (100)	65 (100)
MW (SD)	15,1 (4,56)	15,7 (5,46)
<i>Monat 12</i>		
n (%)	94 (73,4)	51 (78,5)
MW (SD)	14,4 (4,36)	15,6 (5,57)
<i>Veränderung von Baseline zu Monat 12</i>		
n (%) <sup>4)</sup>	126 (98,4) <sup>6)</sup>	63 (96,9) <sup>6)</sup>
LS Mean (SE) <sup>5)</sup>	-0,10 (0,30)	-0,19 (0,41)
LS Mean Difference [95%-KI]; p-Wert <sup>5)</sup>	0,09 [-0,87; 1,06]; 0,85	
<b>Domäne „Fatigue“</b>		
<i>Baseline</i> <sup>3)</sup>		
n (%)	128 (100)	65 (100)
MW (SD)	27,6 (10,0)	27,4 (10,6)
<i>Monat 12</i>		
n (%)	94 (73,4)	51 (78,5)
MW (SD)	23,8 (9,88)	24,9 (9,81)

<b>Studie RESPONSE PBC-40<sup>2)</sup></b>	<b>Seladelpar N = 128</b>	<b>Placebo N = 65</b>
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>4)</sup> LS Mean (SE) <sup>5)</sup>	126 (98,4) <sup>6)</sup> -1,97 (0,67)	63 (96,9) <sup>6)</sup> -1,50 (0,92)
LS Mean Difference [95%-KI]; p-Wert <sup>5)</sup>	-0,47 [-2,66; 1,71]; 0,67	
<b>Domäne „Kognitive Funktion“</b>		
<i>Baseline<sup>3)</sup></i> n (%) MW (SD)	128 (100) 13,2 (5,51)	65 (100) 12,8 (4,88)
<i>Monat 12</i> n (%) MW (SD)	94 (73,4) 12,0 (5,61)	51 (78,5) 12,4 (5,39)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>4)</sup> LS Mean (SE) <sup>5)</sup>	126 (98,4) <sup>6)</sup> -0,55 (0,33)	63 (96,9) <sup>6)</sup> -0,42 (0,44)
LS Mean Difference [95%-KI]; p-Wert <sup>5)</sup>	-0,13 [-1,20; 0,94]; 0,81	
<b>Domäne „Emotionale Funktion“</b>		
<i>Baseline<sup>3)</sup></i> n (%) MW (SD)	128 (100) 7,79 (2,89)	65 (100) 7,72 (3,15)
<i>Monat 12</i> n (%) MW (SD)	94 (73,4) 6,95 (3,02)	51 (78,5) 6,73 (2,84)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>4)</sup> LS Mean (SE) <sup>5)</sup>	126 (98,4) <sup>6)</sup> -0,70 (0,20)	63 (96,9) <sup>6)</sup> -0,95 (0,27)
LS Mean Difference [95%-KI]; p-Wert <sup>5)</sup>	0,25 [-0,40; 0,90]; 0,45	
<b>Domäne „Soziale Funktion“</b>		
<i>Baseline<sup>3)</sup></i> n (%) MW (SD)	128 (100) 22,9 (8,44)	65 (100) 22,2 (8,32)
<i>Monat 12</i> n (%) MW (SD)	94 (73,4) 21,4 (8,53)	51 (78,5) 19,8 (7,10)
<i>Veränderung von Baseline zu Monat 12</i> n (%) <sup>4)</sup> LS Mean (SE) <sup>5)</sup>	126 (98,4) <sup>6)</sup> -0,67 (0,47)	63 (96,9) <sup>6)</sup> -1,72 (0,64)
LS Mean Difference [95%-KI]; p-Wert <sup>5)</sup>	1,05 [-0,47; 2,56]; 0,17	

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> Erhebung der gesundheitsbezogenen Lebensqualität bei Teilnehmenden mit PBC, bestehend aus 40 Items in 6 Domänen. Je nach Anzahl der Items in den einzelnen Domänen ergeben sich unterschiedliche Wertebereiche für die Scores der Domänen, die von 3 bis 55 Punkten reichen. Höhere Punktzahlen spiegeln eine Verschlechterung wider, während niedrigere Punktzahlen eine Verbesserung bedeuten.

<sup>3)</sup> Baseline war definiert als arithmetisches Mittel aller verfügbaren Messungen bei Screening, während der Run-in-Phase (ca. 2 Wochen vor Tag 1), Tag 1 und ggf. ungeplanten Erhebungen vor oder an Tag 1.

- <sup>4)</sup> Anzahl an Personen, die sowohl einen Baseline-Wert als auch mindestens einen Wert nach Baseline aufwiesen und somit in das MMRM eingeschlossen wurden.
- <sup>5)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-Werten, Randomisierungsstratum zu Studienbeginn (ALP-Wert < 350 U/L vs. ALP-Wert  $\geq$  350 U/L und Pruritus NRS < 4 vs. Pruritus NRS  $\geq$  4), Behandlungsgruppe, Woche und die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline des PBC-40-Gesamtscores wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als Term in die Modellanalyse für die Domäne „Soziale Funktion“ aufgenommen.
- <sup>6)</sup> Prozentangaben aus eigener Berechnung.

Abkürzungen: ALP: Alkalische Phosphatase; ITT: Intention to Treat; KI: Konfidenzintervall; LS: Least Squares; MMRM: Gemischtes Modell mit Messwiederholungen; MW: Mittelwert; PBC: Primär biliäre Cholangitis; SD: Standardabweichung; SMD: Standardisierte Mittelwertdifferenz.

Für die post-hoc Subgruppenanalyse der kognitiven Domäne zeigte sich zu Monat 12 eine signifikante Interaktion für das Subgruppenmerkmal Region (p = 0,04). Bei Personen aus Europa zeigte sich ein signifikanter Vorteil für Seladelpar (LS Mean Difference [95%-KI]: -1,88 [-3,67; -0,09]; p = 0,04). Bei Personen aus Nordamerika (LS Mean Difference [95%-KI]: 1,37 [-0,55; 3,30]; p = 0,16) sowie aus der restlichen Welt (LS Mean Difference [95%-KI]: 0,44 [-1,52; 2,40]; p = 0,65) zeigte sich kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen. Auch in der Fatigue-Domäne zeigte sich eine signifikante Interaktion für das Merkmal „Region“ (p = 0,03), auch hier mit einem signifikanten Vorteil für Seladelpar bei Personen aus Europa (LS Mean Difference [95%-KI]: -5,06 [-8,85; -1,28]; p = 0,01). Bei Personen aus Nordamerika (LS Mean Difference [95%-KI]: 0,10 [-3,93; 4,13]; p = 0,96) und der restlichen Welt (LS Mean Difference [95%-KI]: 1,72 [-2,08; 5,53]; p = 0,37) zeigte sich kein statistisch signifikanter Effekt zwischen den Behandlungsarmen.

In der Gesamtschau wurde für das Merkmal „Region“ ein signifikanter Interaktionsterm lediglich in zwei der PBC-40-Domänen beobachtet. In den weiteren PBC-40-Domänen sowie bei weiteren in der Nutzenbewertung dargestellten Endpunkten zeigte sich für das Merkmal „Region“ keine Effektmodifikation.

Bezüglich der Herkunftsregion der Teilnehmenden zeigten sich trotz Randomisierung Ungleichgewichte: Im Seladelpar-Arm (39 %) wurden knapp 19 % mehr Personen aus Nordamerika eingeschlossen als im Placebo-Arm (20 %). Aus Lateinamerika wurden im Seladelpar-Arm 19 % eingeschlossen, etwa 10 % weniger als im Placebo-Arm (29 %). Inwieweit sich diese unterschiedliche regionale Verteilung trotz Randomisierung auf die Ergebnisse der Subgruppenanalyse auswirkt, lässt sich nicht abschließend beurteilen.

### 3.5 Sicherheit

Der pU legt für die Gesamtraten der UE in Modul 4 zusätzliche Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Es ist jedoch unklar, welche erkrankungsbezogene Ereignisse herausgerechnet wurden. Daher werden diese Analysen nicht dargestellt.

Für Teilnehmende, die nach Monat 12 in die Langzeitnachbeobachtungsstudie ASSURE wechselten, entfiel die reguläre Sicherheitsnachbeobachtung im Rahmen der RESPONSE-Studie. Etwa 83 % in beiden Studienarmen gingen in die Langzeitnachbeobachtungsstudie (LTE) über (siehe Tabelle 9). Eine Person im Seladelpar-Arm beendete zwar die Behandlungsphase der RESPONSE-Studie, schloss jedoch die Sicherheitsnachbeobachtung nicht ab und wechselte auch nicht in die LTE ASSURE.

Die mediane Beobachtungsdauer für die Sicherheitsüberwachung lag im Seladelpar-Arm bei 52,4 Wochen (min: 4,4; max: 58,3) und im Placebo-Arm bei 52,1 Wochen (min: 5,6; max: 58,1).

**Tabelle 17: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase, Studie RESPONSE, Sicherheitspopulation, Datenschnitt: 11.08.2023<sup>1)</sup>**

<b>Studie RESPONSE Zusammenfassung der UE Personen mit mindestens einem ...</b>	<b>Seladelpar N = 128 n (%)</b>	<b>Placebo N = 65 n (%)</b>	<b>RR [95%-KI]<sup>2)</sup>; p-Wert</b>
UE (ergänzend dargestellt)	111 (86,7)	55 (84,6)	-
UE CTCAE-Grad $\geq 3$	14 (10,9)	5 (7,7)	1,42 [0,54; 3,78]; 0,48
SUE	9 (7,0)	4 (6,2)	1,14 [0,37; 3,57]; 0,82
UE, das zum Abbruch der Studienmedikation führte <sup>3)</sup>	4 (3,1)	3 (4,6)	0,68 [0,16; 2,94]; 0,60

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> Das RR wurde post-hoc berechnet anhand der Mantel-Haenszel-Methode. Die 95%-Konfidenzintervalle und p-Werte wurden unter Verwendung der Normalapproximation (Wald-Test) berechnet. Es ist nicht eindeutig ersichtlich, ob das RR stratifiziert nach den Randomisierungsfaktoren berechnet wurde. Ein RR < 1 bedeutet einen Vorteil für die Behandlung mit Seladelpar.

<sup>3)</sup> Die Studienteilnehmenden erhielten die Studienmedikation u. a. bis zum Auftreten inakzeptabler UE, klinischer PBC-Ereignisse, oder der Rücknahme der Einverständniserklärung, je nachdem, was früher eintrat. Diese möglichen Therapieabbruchgründe, die vor einem potenziellen Abbruch aufgrund von UE auftreten können, stellen somit ein potenzielles konkurrierendes Ereignis dar. Im Studienverlauf bis zum Studienende (11.08.2023) trat jedoch lediglich ein einziges klinisches PBC-Ereignis auf, und dies erst 7 Tage nach Abschluss der Behandlungsphase während der Sicherheitsnachbeobachtung. Auch die Anzahl der Studienabbrüche aufgrund einer Rücknahme der Einverständniserklärung war in beiden Studienarmen gering (jeweils 3,1 %). Daher wird nicht davon ausgegangen, dass die Ergebnissicherheit oder die Interpretierbarkeit der Ergebnisse in relevantem Ausmaß beeinträchtigt wurden.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; KI: Konfidenzintervall; PBC: Primär biliäre Cholangitis; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

### Unerwünschte Ereignisse

Für die UE nach SOC und PT liegen keine Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen vor.

**Tabelle 18: UE mit Inzidenz  $\geq 10$  % oder bei  $\geq 10$  Patienten sowie mindestens 1 % in mindestens einem Studienarm nach Systemorganklasse und Preferred Term in der Studie RESPONSE, Datenschnitt: 11.08.2023<sup>1)</sup>**

<b>Studie RESPONSE MedDRA-Systemorganklasse<sup>2)</sup> Preferred Term</b>	<b>Seladelpar N = 128 n (%)</b>	<b>Placebo N = 65 n (%)</b>	<b>RR [95%-KI]<sup>3)</sup>; p-Wert</b>
<b>Erkrankungen des Blutes und des Lymphsystems</b>	<b>15 (11,7)</b>	<b>3 (4,6)</b>	<b>2,54 [0,76; 8,46]; 0,13</b>
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>41 (32,0)</b>	<b>24 (36,9)</b>	<b>0,87 [0,58; 1,30]; 0,49</b>
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>23 (18,0)</b>	<b>13 (20,0)</b>	<b>0,90 [0,49; 1,66]; 0,73</b>
<b>Infektionen und parasitäre Erkrankungen</b>	<b>58 (45,3)</b>	<b>35 (53,8)</b>	<b>0,84 [0,63; 1,13]; 0,25</b>
COVID-19	23 (18,0)	10 (15,4)	1,17 [0,59; 2,30]; 0,65
<b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>	<b>17 (13,3)</b>	<b>4 (6,2)</b>	<b>2,16 [0,76; 6,15]; 0,15</b>
<b>Untersuchungen</b>	<b>18 (14,1)</b>	<b>6 (9,2)</b>	<b>1,52 [0,64; 3,65]; 0,35</b>

Studie RESPONSE MedDRA-Systemorganklasse <sup>2)</sup> Preferred Term	Seladelpar N = 128 n (%)	Placebo N = 65 n (%)	RR [95%-KI] <sup>3)</sup> ; p-Wert
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>15 (11,7)</b>	<b>10 (15,4)</b>	<b>0,76 [0,36; 1,60]; 0,47</b>
<b>Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen</b>	<b>31 (24,2)</b>	<b>18 (27,7)</b>	<b>0,87 [0,53; 1,44]; 0,60</b>
<b>Erkrankungen des Nervensystems</b>	<b>22 (17,2)</b>	<b>9 (13,8)</b>	<b>1,24 [0,61; 2,54]; 0,55</b>
Kopfschmerzen	10 (7,8)	2 (3,1)	2,54 [0,57; 11,3]; 0,22
<b>Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>	<b>12 (9,4)</b>	<b>5 (7,7)</b>	<b>1,22 [0,45; 3,31], 0,70</b>
<b>Erkrankungen der Haut und des Unterhautgewebes</b>	<b>20 (15,6)</b>	<b>15 (23,1)</b>	<b>0,68 [0,37; 1,23] 0,20</b>
Pruritus	6 (4,7)	10 (15,4)	0,30 [0,12; 0,80] 0,02

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> MedDRA-Version 26.0; CTCAE-Version 4.03.

<sup>3)</sup> Das RR wurde post-hoc berechnet anhand der Mantel-Haenszel-Methode. Die 95%-Konfidenzintervalle und p-Werte wurden unter Verwendung der Normalapproximation (Wald-Test) berechnet. Es ist nicht eindeutig ersichtlich, ob das RR stratifiziert nach den Randomisierungsfaktoren berechnet wurde. Ein RR < 1 bedeutet einen Vorteil für die Behandlung mit Seladelpar.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; RR: Relatives Risiko; UE: Unerwünschtes Ereignis.

### Schwere UE

In der Studie RESPONSE traten bei den schweren UE keine SOC und PT mit mindestens 5 % Inzidenz in einer der Behandlungsgruppen auf.

### Schwerwiegende UE

In der Studie RESPONSE traten bei den SUE keine SOC und PT mit mindestens 5 % Inzidenz in einer der Behandlungsgruppen auf.

### UE von besonderem Interesse (AESI)

*Tabelle 19: UE von besonderem Interesse<sup>1)</sup> nach Systemorganklasse und Preferred Term in der Studie RESPONSE, Sicherheitspopulation, Datenschnitt: 11.08.2023<sup>2)</sup>*

Studie RESPONSE MedDRA-Systemorganklasse Preferred Term	Seladelpar N = 128 n (%)	Placebo N = 65 n (%)	RR [95%-KI] <sup>3)</sup> ; p-Wert
<b>Kardiovaskuläre Toxizität</b>			
UE (unabhängig vom Schweregrad), n (%)	13 (10,2)	5 (7,7)	1,32 [0,49; 3,54]; 0,58
Schwere UE (Grad ≥ 3), n (%)	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
SUE	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
<b>Herzrhythmusstörungen (SMQ)</b>			
UE (unabhängig vom Schweregrad), n (%)	3 (2,3)	2 (3,1)	0,76 [0,13; 4,45]; 0,76
Schwere UE (Grad ≥ 3), n (%)	0 (0,0)	0 (0,0)	n. b.
SUE	0 (0,0)	0 (0,0)	n. b.

<b>Studie RESPONSE</b> <b>MedDRA-Systemorganklasse</b> Preferred Term	<b>Seladelpar</b> <b>N = 128</b> n (%)	<b>Placebo</b> <b>N = 65</b> n (%)	<b>RR [95%-KI]<sup>3)</sup>; p-Wert</b>
<b>Herzinsuffizienz (SMQ)</b>			
UE (unabhängig vom Schweregrad), n (%)	3 (2,3)	1 (1,5)	1,52 [0,16; 14,4]; 0,71
Schwere UE (Grad $\geq$ 3), n (%)	0 (0,0)	0 (0,0)	n. b.
SUE	0 (0,0)	0 (0,0)	n. b.
<b>Kardiomyopathie (SMQ)</b>			
UE (unabhängig vom Schweregrad), n (%)	5 (3,9)	0 (0,0)	5,63 [0,32; 100,2]; 0,24
Schwere UE (Grad $\geq$ 3), n (%)	0 (0,0)	0 (0,0)	n. b.
SUE	0 (0,0)	0 (0,0)	n. b.
<b>Ischämische Herzkrankheit (SMQ)</b>			
UE (unabhängig vom Schweregrad), n (%)	5 (3,9)	2 (3,1)	1,27 [0,25; 6,37]; 0,77
Schwere UE (Grad $\geq$ 3), n (%)	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
SUE	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
<b>Lebertoxizität</b>			
UE (unabhängig vom Schweregrad), n (%)	8 (6,3)	6 (9,2)	0,68 [0,25; 1,87]; 0,45
Schwere UE (Grad $\geq$ 3), n (%)	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
SUE	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
<b>Muskeltoxizität</b>			
UE (unabhängig vom Schweregrad), n (%)	8 (6,3)	5 (7,7)	0,81 [0,28; 2,38]; 0,71
Schwere UE (Grad $\geq$ 3), n (%)	0 (0,0)	1 (1,5)	0,17 [0,01; 4,13]; 0,28
SUE	0 (0,0)	0 (0,0)	n. b.
<b>Renale Toxizität</b>			
UE (unabhängig vom Schweregrad), n (%)	0 (0,0)	0 (0,0)	n. b.
Schwere UE (Grad $\geq$ 3), n (%)	0 (0,0)	0 (0,0)	n. b.
SUE	0 (0,0)	0 (0,0)	n. b.
<b>Pankreasbedingte Toxizität</b>			
UE (unabhängig vom Schweregrad), n (%)	2 (1,6)	1 (1,5)	1,02 [0,09; 10,9]; 0,99
Schwere UE (Grad $\geq$ 3), n (%)	0 (0,0)	0 (0,0)	n. b.
SUE	0 (0,0)	0 (0,0)	n. b.
<b>Pruritus-assoziiert</b>			
UE (unabhängig vom Schweregrad), n (%)	7 (5,5)	10 (15,4)	0,36 [0,14; 0,89]; 0,03
Schwere UE (Grad $\geq$ 3), n (%)	0 (0,0)	1 (1,5)	0,17 [0,01; 4,13]; 0,28
SUE	0 (0,0)	0 (0,0)	n. b.
<b>Frakturen<sup>4)</sup></b>			
UE (unabhängig vom Schweregrad), n (%)	7 (5,5)	0 (0,0)	7,67 [0,45; 132,3]; 0,16
Schwere UE (Grad $\geq$ 3), n (%)	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79
SUE	1 (0,8)	0 (0,0)	1,53 [0,06; 37,2]; 0,79

- <sup>1)</sup> In den Studienunterlagen des Dossiers wird der Begriff „UE von Interesse“ verwendet.
- <sup>2)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.
- <sup>3)</sup> Das RR wurde post-hoc berechnet anhand der Mantel-Haenszel-Methode. Die 95%-Konfidenzintervalle und p-Werte wurden unter Verwendung der Normalapproximation (Wald-Test) berechnet. Es ist nicht eindeutig ersichtlich, ob das RR stratifiziert nach den Randomisierungsfaktoren berechnet wurde. Ein  $RR < 1$  bedeutet einen Vorteil für die Behandlung mit Seladelpar.
- <sup>4)</sup> Post-hoc in Modul 4 als „UE von Interesse“ berichtet.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; n. b.: nicht bewertbar; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## 4 Diskussion der Methodik und Ergebnisse

### 4.1 Zulassungsstatus und Zulassungspopulation von Seladelpar

Seladelpar (Seladelpar Gilead®) ist zugelassen zur Behandlung der primär biliären Cholangitis (PBC) in Kombination mit Ursodeoxycholsäure (UDCA) bei Erwachsenen, die nicht ausreichend auf UDCA allein ansprechen, oder als Monotherapie bei Patientinnen und Patienten, die UDCA nicht vertragen. Gemäß Fachinformation beträgt die empfohlene Dosierung 10 mg einmal täglich.

Grundlage der Nutzenbewertung ist die zulassungsbegründende Studie RESPONSE, in die Erwachsene im Alter zwischen 18 und 75 Jahren mit nachgewiesener PBC eingeschlossen werden konnten, die seit mindestens 12 Monaten mit UDCA behandelt wurden oder UDCA nicht vertragen. Die Studie wurde an insgesamt 90 Studienzentren in 24 Ländern auf den Kontinenten Amerika, Europa, Asien und Australien durchgeführt, darunter auch in Deutschland. Etwa 39 % der Teilnehmenden waren aus dem EMEA-Raum (Europa, Naher Osten und Afrika), es ist jedoch unklar wie viele davon aus Europa kamen und wie viele Studienzentren in Deutschland beteiligt waren. Die Zusammenfassung von Europa, dem nahen Osten und Afrika zu einer Kategorie wird zudem als nicht differenziert genug erachtet, da sich die Versorgung in den Regionen stark unterscheiden kann. An der Studie nahmen größtenteils Frauen teil, mit einem Anteil von ca. 95 %, was der für dieses Krankheitsbild charakteristischen Geschlechterverteilung weitgehend entspricht. Lediglich etwa 6 % im Seladelpar-Arm und etwa 6 % im Placebo-Arm waren UDCA-intolerant. Alle eingeschlossenen Personen wurden bereits mit UDCA behandelt. Es ist unklar, inwiefern die eingeschlossenen Personen, die gegenüber UDCA tolerant waren, ein unzureichendes Ansprechen auf UDCA entsprechend des zugelassenen Anwendungsgebiets zeigten. Den Studienunterlagen konnte keine eindeutige Definition oder spezifische Kriterien entnommen werden, anhand derer ein unzureichendes Ansprechen auf UDCA operationalisiert wurde. Personen mit einer fortgeschrittenen PBC im Sinne der Rotterdam-Kriterien (Albumin < LLN UND Gesamt-Bilirubin > 1,0 x ULN) wurden explizit von der Studie ausgeschlossen. Die Studie bildet demnach nicht alle Stadien einer PBC ab und ist diesbezüglich gegenüber der Zulassungspopulation eingeschränkt. Mit Protokollamendment 4 (09.02.2022) wurden in Rücksprache mit der FDA die Einschlusskriterien für die geschätzte glomeruläre Filtrationsrate (eGFR) von > 60 mL/min/1,73 m<sup>2</sup> (normal bis leicht eingeschränkte Nierenfunktion) auf ≥ 45 mL/min/1,73 m<sup>2</sup> (moderat eingeschränkte Nierenfunktion) angepasst. Somit konnten erst mit der finalen Protokollversion auch Personen mit moderat eingeschränkter Nierenfunktion eingeschlossen werden. Dies spiegelt sich in den eGFR-Werten zu Baseline wider: 75 % der Teilnehmenden hatten eine eGFR über ca. 82 ml/min/1,73 m<sup>2</sup> (also eine normale bis sehr gute Nierenfunktion). Der Median lag in beiden Studienarmen bei knapp 96 mL/min/1,73 m<sup>2</sup>. Wie viele der Teilnehmenden konkret einen eGFR-Wert unter 60 mL/min/1,73 m<sup>2</sup> aufwiesen, geht aus den vorliegenden Daten nicht hervor. Insgesamt hatte jedoch die Mehrheit der Teilnehmenden keine oder nur eine leichte Einschränkung. Daher könnte die Studienpopulation auch in Bezug auf die Nierenfunktion gegenüber der Zulassungspopulation eingeschränkt sein.

Laut EPAR handelt es sich bei Seladelpar um einen neuen Wirkstoff, der am 20. Februar 2025 im vorliegenden Anwendungsgebiet zugelassen wurde. Im Rahmen der bedingten Zulassung werden Ergebnisse vor allem hinsichtlich langfristiger klinischer Endpunkte der randomisiert-kontrollierten Studie „AFFIRM“ mit einer Studiendauer von 156 Wochen gefordert, in die erwachsene Patientinnen und Patienten mit primär biliärer Cholangitis (PBC) und kompensierter Zirrhose eingeschlossen werden. Die Vorlage dieser Ergebnisse ist bis August 2030 vorgesehen [2,7].

## 4.2 Design und Methodik der Studie

Die Nutzenbewertung stützt sich auf die pivotale Studie RESPONSE, eine multizentrische, doppelblinde, randomisierte, placebokontrollierte Phase-III-Parallelgruppenstudie zur Untersuchung der Wirksamkeit, Verträglichkeit und Pharmakokinetik von Seladelpar bei Erwachsenen mit PBC. Auf eine bis zu 3-wöchige Screening-Phase folgte eine 2-wöchige Run-in-Periode und anschließend der 12-monatige Behandlungszeitraum. Der Zeitraum für die Sicherheitsnachbeobachtung war 2 Wochen (+ 3 Tage). Während Teilnehmende mit UDCA-Unverträglichkeit Seladelpar als Monotherapie erhielten (6 % in beiden Studienarmen), wurde Seladelpar bei Personen, die zu Baseline mit UDCA behandelt wurden, in Kombination mit UDCA verabreicht. Eine stabile UDCA-Dosis sollte seit 3 Monaten vor Screening bestehen bzw. sollte bei Unverträglichkeit die letzte Dosis zum Zeitpunkt des Screenings mehr als 3 Monate zurückliegen. Nach Abschluss des 12-monatigen Behandlungszeitraums hatten die Teilnehmenden die Möglichkeit, an einer offenen Langzeitnachbeobachtungsstudie (ASSURE) teilzunehmen. Für Patientinnen und Patienten, die in die Studie ASSURE übergangen, entfiel die reguläre Sicherheitsnachbeobachtung der Studie RESPONSE.

Insgesamt wurden 193 Patientinnen und Patienten randomisiert, davon 128 in den Seladelpar-Arm und 65 in den Placebo-Arm. Die letzte Visite des letzten Teilnehmenden erfolgte am 11.08.2023, womit die Studie abgeschlossen ist. Der Zeitpunkt des Datenschnitts geht nicht eindeutig aus den Studienunterlagen hervor. In Modul 4 wird der 11.08.2023 als Datenschnitt angegeben, was dem Studienenddatum entspricht. Im Anschluss an die 12-monatige Behandlungsphase der Studie RESPONSE wechselten gemäß Studienbericht 160 (bzw. gemäß EPAR 159) der ursprünglich randomisierten Teilnehmenden in die offene, einarmige Extensionsstudie ASSURE. Da bei der Studie ASSURE ein einarmiges Studiendesign angewendet wird und parallel ein geeignetes randomisiertes, kontrolliertes Studiendesign mit 1-jähriger Beobachtungsdauer vorliegt, wird die Studie ASSURE für die Nutzenbewertung nicht berücksichtigt.

Studienübergreifend brachen knapp 10 % die Studie vorzeitig ab, dies entspricht 9 % im Seladelpar-Arm sowie 12 % im Placebo-Arm. Knapp 91 % der Teilnehmenden schlossen die Studienbehandlung ab, wobei dies auf etwa 92 % im Seladelpar-Arm und rund 88 % im Placebo-Arm zutraf. Im Seladelpar-Arm waren mit jeweils ca. 3 % die häufigsten Gründe für den Abbruch der Einnahme der Studienmedikation unerwünschte Ereignisse sowie der Entzug der Einverständniserklärung. Im Placebo-Arm war der häufigste Grund das Auftreten von UE (ca. 6 %). In beiden Studienarmen kam es bei etwa der Hälfte der Teilnehmenden zu einer signifikanten Protokollverletzung. Anhand der dokumentierten Protokollverletzungen ergeben sich jedoch keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

Primäre Endpunkte waren das „kombinierte biochemische Ansprechen“ zu Monat 12 sowie unerwünschte Ereignisse. Zu den sekundären Endpunkten gehörten Pruritus (u. a. „Pruritus NRS“, „5-D Itch“ und „PGI-S“) sowie die Lebensqualität („PBC-40“). Es werden Auswertungen anhand der ITT-Population (Seladelpar: N = 128; Placebo: N = 65) bzw. Sicherheitspopulation (übereinstimmend) dargestellt.

Es erfolgten 2 Protokollamendments (Amendment 3 und 4) nach Einschluss der ersten Person. Durch die Protokolländerungen werden keine bedeutenden Auswirkungen auf die Ergebnisse angenommen.

## Studienpopulation

Die Teilnehmenden waren mehrheitlich weiblich (ca. 95 %) und im Mittel knapp 57 Jahre.

Gemäß Einschlusskriterien waren Personen ab einem Alter von 18 Jahren für die Studie geeignet. Die tatsächlich eingeschlossenen Teilnehmenden im Seladelpar-Arm waren mindestens 28 Jahre alt, im Placebo-Arm mindestens 33 Jahre. Zudem waren 95 % der eingeschlossenen Personen älter als 37 (Seladelpar-Arm) bzw. 39 Jahre (Placebo-Arm).

Bezüglich der Herkunftsregion der Teilnehmenden zeigten sich trotz Randomisierung Ungleichgewichte: Im Seladelpar-Arm (39 %) wurden knapp 19 % mehr Personen aus Nordamerika eingeschlossen als im Placebo-Arm (20%) und im Seldalpar-Arm (19 %) wurden etwa 10 % weniger aus Lateinamerika eingeschlossen als im Placebo-Arm (29 %). Im Seladelpar-Arm lag zudem der Anteil der Teilnehmenden mit hispanischer oder latein-amerikanischer ethnischer Zugehörigkeit um knapp 19 % niedriger (23 %) als im Placebo-Arm (42 %). Laut EPAR ergaben angeforderte Analysen zu den Unterschieden hinsichtlich Region und ethnischer Zugehörigkeit, dass dies keine relevanten Auswirkungen auf die Ergebnisse zur Folge hatte.

Grundsätzlich befand sich der Großteil der Teilnehmenden in beiden Studienarmen (86 %) zu Baseline in einem „milden“ Krankheitsstadium gemäß Rotterdam-Kriterien, also zeigten ein normales Gesamtbilirubin- und normales Albumin-Level. Nur insgesamt 14 % zeigten ein „mäßig fortgeschrittenes“ Stadium (entweder abnormes Albumin oder abnormes Gesamtbilirubin). Keine Person wurde als „fortgeschritten“ eingestuft (sowohl abnormes Albumin als auch abnormes Gesamtbilirubin) und nur knapp 14 % aller Teilnehmenden wurden zu Baseline als zirrhotisch eingestuft. Im Seladelpar-Arm befanden sich fast 10 % mehr Patientinnen und Patienten mit einem moderat fortgeschrittenen Rotterdam-Krankheitsstadium. Die übrigen Werte zur Leberfunktion- und -struktur sowie die krankheits-relevanten Laborparameter waren jedoch vergleichbar.

Auch hinsichtlich des Juckreizes zeigten die meisten Teilnehmenden zu Studienbeginn eher gering ausgeprägte Symptome. Etwa 63 % der Studienpopulation wiesen einen Baseline-Wert auf der Pruritus NRS von unter 4 auf. Der Median lag im Seladelpar-Arm bei 2,8 und im Placebo-Arm bei 1,8. Darüber hinaus hatten 75 % der Teilnehmenden im Seladelpar-Arm zu Studienbeginn einen Pruritus-NRS-Wert von  $\leq 5,3$  und im Placebo-Arm von  $\leq 5,6$ . Zudem lag bei 25 % der Teilnehmenden im Seladelpar-Arm der Baseline-Wert bei  $\leq 0,1$  und im Placebo-Arm bei  $\leq 0,2$ ; konkret hatten 30 Teilnehmende im Seladelpar-Arm und 15 im Placebo-Arm einen Wert von genau 0 und damit keinen Juckreiz zu Studienbeginn. Angesichts der Skalenspannweite der Pruritus NRS (0–10) und der vom pU vorgegebenen Einteilung, wonach ein Pruritus-NRS-Wert  $\geq 4$  als moderater bis schwerer Juckreiz gilt, hatten die meisten Studienteilnehmenden entweder keinen oder lediglich einen leichten bis moderaten Juckreiz.

Die Verteilung der Begleitmedikation und therapeutischen Begleitinterventionen war in beiden Gruppen weitgehend ähnlich, mit Ausnahme der lipidmodifizierenden Substanzen: Diese wurden im Placebo-Arm von 52,3 % und im Seladelpar-Arm von 32,8 % der Teilnehmenden eingenommen, was einem Unterschied von etwa 20 % entspricht. Vorliegend wird nicht davon ausgegangen, dass diese Imbalance relevante Auswirkungen auf die Ergebnisse hat.

Insgesamt wird das Verzerrungspotential auf Studienebene als niedrig eingeschätzt.

### 4.3 Mortalität

Todesfälle jeglicher Ursache wurden im Rahmen der Sicherheit erhoben. In der Studie RESPONSE sind bis zum Ende der Studie vom 11.08.2023 keine Todesfälle aufgetreten.

### 4.4 Morbidität

Im Hinblick auf die Morbidität werden zwei Endpunkte zur Erhebung des Juckreizes („Pruritus NRS“ und „5-D Itch“) sowie der „MELD-Score  $\geq$  15 bei mindestens 2 aufeinanderfolgenden Visiten“ herangezogen.

Zwei weitere Endpunkte zur Erhebung des Juckreizes werden im Anhang dargestellt: der Endpunkt „PGI-S“ aufgrund von Mehrfacherfassung und der Endpunkt „PGI-C“ (aufgrund von unklarer Validität und Mehrfacherfassung) aus Transparenzgründen. Die grundsätzlich patientenrelevanten klinischen PBC-Ereignisse „Lebertransplantation“, „behandlungsbedürftiger Aszites“ und „Hospitalisierung“ werden aufgrund ihrer unklaren Validität im Anhang dargestellt. Aus Transparenzgründen wird auch der primäre Endpunkt „kombiniertes biochemisches Ansprechen“ im Anhang dargestellt.

Bei dem vorliegenden Anwendungsgebiet handelt es sich um eine langsam progrediente Erkrankung, bei der die Auswertung unter Berücksichtigung eines längeren Zeitverlaufs gegenüber einer Auswertung zu nur einem Erhebungszeitpunkt bevorzugt wird. Daher werden die Auswertungen der kontinuierlichen Daten mittels MMRM grundsätzlich als adäquat angesehen und für die Nutzenbewertung herangezogen. Für die „5-D Itch“ erfolgten MMRM-Analysen post-hoc, während diese für die „Pruritus NRS“ präspezifiziert waren. Zusätzlich werden für die „Pruritus NRS“ präspezifizierte Responderanalysen dargestellt. Aufgrund von Unklarheiten im Erhebungsvorgehen für die „Pruritus NRS“ nach Monat 6 und der damit verbundenen unklaren Validität und dem hohen Verzerrungspotenzial der Analysen nach Monat 6 werden vorrangig die Ergebnisse zu Monat 6 dargestellt; die Ergebnisse zu Monat 12 werden ergänzend berichtet.

In Bezug auf die in Modul 4 berichteten Subgruppenanalysen zeigen sich keine konsistenten Effektmodifikationen über mehrere Auswertungszeitpunkte oder Domänen innerhalb eines Endpunktes oder über mehrere Endpunkte derselben Symptomkategorie hinweg. Aus diesem Grund werden die Ergebnisse der Subgruppenanalysen in der Nutzenbewertung nicht weiter diskutiert.

#### **Pruritus**

Bei der Interpretation der Ergebnisse zu den Juckreiz-bezogenen Endpunkten ist zu berücksichtigen, dass einige Teilnehmende Juckreiz-stillende Begleitmedikation erhielten: Unter anderem erhielten etwa 7 % der Teilnehmenden Cholestyramin und knapp 3 % Rifampicin, welche laut EPAR jeweils bei cholestatischem Juckreiz angewendet werden. Die Anzahl der Personen, die diese Medikamente erhielten, war zwischen den Studienarmen weitgehend vergleichbar und wirken sich vorliegend nicht auf das Verzerrungspotenzial aus.

#### ***Pruritus NRS***

Die mediane Beobachtungsdauer für den Endpunkt „Pruritus NRS“ lag in beiden Studienarmen bei 46,6 Wochen (jeweils min: 0,1; max: 46,6).

Zu Monat 6 zeigte sich ein statistisch signifikanter Unterschied zugunsten von Seladelpar gegenüber Placebo (LS Mean Difference [95%-KI]; p-Wert: -0,90 (-1,35; -0,45); 0,0001). Auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz; post-hoc berichtet) erscheint in dieser MMRM-Analyse ein klinisch relevanter Effekt plausibel (SMD [95%-KI]: -0,58 [-0,89; -0,28]).

Für die Responderanalyse zu Monat 6 liegen keine Effektschätzer oder p-Werte vor. Rein numerisch zeigen sich im Seladelpar-Arm mehr Response-Ereignisse, also ein größerer Anteil an Personen mit einer Reduktion um  $\geq 4$  Punkte oder auf 0 zu Monat 6 im Vergleich zum Placebo-Arm. Während im Seladelpar-Arm 28 Personen (21,9% bezogen auf die ITT-Population) als Responder klassifiziert werden, sind es im Placebo-Arm 10 Personen (15,4% bezogen auf die ITT-Population).

Der Endpunkt wird im vorliegenden Anwendungsgebiet grundsätzlich als valide eingestuft. Allerdings ergeben sich potenzielle Einschränkungen der Validität der Auswertungen aufgrund der Methodik, wie den Zeitpunkten der Studienvisiten Werte für die Pruritus NRS zugeordnet werden. Für die Auswertung Post-Baseline verwendet der pU jeweils den Mittelwert eines 7-Tage-Intervalls, der den Zeitpunkten der Studienvisiten zugeordnet wird. Durch diese Vorgehensweise ist nicht nachvollziehbar, wie viele der tatsächlichen Erhebungen fehlen. Zudem werden die Mittelwerte der 7-Tage-Intervalle bereits berechnet, sobald mindestens ein Wert vorliegt. Angaben zur Vollständigkeit der e-Tagebuch-Einträge liegen nicht vor. Daher bleibt unklar, wie viele Daten tatsächlich in die Analyse eingeflossen sind, und es kann nicht ausgeschlossen werden, dass fehlende Werte zwischen den Studienarmen ungleich verteilt waren. Aufgrund der beschriebenen Limitationen wird das Verzerrungspotenzial für die Analysen zu Monat 6 als „unklar“ eingestuft.

### **5-D-Itch-Scale**

Die mediane Beobachtungsdauer für den Endpunkt „5-D Itch“ lag in beiden Studienarmen bei 44 Wochen (für Seladelpar min: 4,1; max: 46; für Placebo min: 0,1; max: 45,1). Bei der Interpretation der Ergebnisse zu Monat 12 ist somit zu berücksichtigen, dass keine tatsächlichen Werte zu Woche 52 vorlagen.

Für den Endpunkt „5-D Itch“ zeigte sich für den Gesamtscore zu Monat 12 ein statistisch signifikanter Unterschied zugunsten von Seladelpar gegenüber Placebo (LS Mean Difference [95%-KI]; p-Wert: -2,31 [-3,40; -1,23];  $< 0,0001$ ). Auf Basis von Hedges' g (post-hoc berichtet), erscheint in dieser MMRM-Analyse ein klinisch relevanter Effekt plausibel (SMD [95%-KI]: -0,62 [-0,93; -0,31]). Auch in den einzelnen Domänen zeigte sich jeweils ein statistisch signifikanter Unterschied zugunsten von Seladelpar im Vergleich zu Placebo, wobei gemäß des post-hoc berichteten Hedges' g nur für die Domäne „Beeinträchtigung“ ein klinisch relevanter Effekt plausibel erscheint.

Die 5-D Itch wird im vorliegenden Anwendungsgebiet grundsätzlich als valide eingeschätzt. Der Gesamtscore ist umfangreich validiert. Aufgrund der in Kapitel 2.3.2 beschriebenen Unsicherheiten in Bezug auf die Domäne „Verlauf“ berichtet der pU einen nicht-validierten, modifizierten Gesamtscore, aus dem die Domäne „Verlauf“ herausgerechnet wird. Der modifizierte Gesamtscore wird aufgrund fehlender Validierung nicht berücksichtigt. Die Argumentation des pU kann jedoch grundsätzlich nachvollzogen werden, weshalb zusätzlich zum validierten Gesamtscore die einzelnen Domänen herangezogen werden. Domänenübergreifend ergibt sich hinsichtlich der Ergebnisse ein konsistentes Bild, das mit den Resultaten des Gesamtscores übereinstimmt. Daher ist nicht davon auszugehen, dass die Unsicherheiten in der Verlaufs-Domäne die Aussagesicherheit der Ergebnisse des Gesamtscores beeinträchtigen.

Die Zuordnung der 5-D-Itch-Werte zu den Zeitpunkten der Studienvisiten erfolgte anhand festgelegter Zieltage, wobei aus jeweils präspezifizierten Zeitfenstern der dem Zieltag am nächsten liegende Wert ausgewählt wurde; bei gleichem Abstand wurde der später erhobene Wert verwendet. Aufgrund der teilweise sehr breiten Zeitfenster um die Zieltage könnte die Parallelität der Erhebungen zwischen den Teilnehmenden und somit auch den Studienarmen möglicherweise nicht mehr gewährleistet sein. Daten zur Vollständigkeit der Erhebungen im Rahmen des e-Tagebuchs liegen nicht vor. Dadurch kann nicht ausgeschlossen werden, dass fehlende Werte zwischen den Studienarmen ungleich verteilt waren. Zudem lagen die Rücklaufquoten zu Monat 12 im Seladelpar-Arm bei nur 75 % und im Placebo-Arm bei nur rund 71 %. Aufgrund der beschriebenen Limitationen wird das Verzerrungspotenzial als hoch eingestuft.

#### **MELD-Score $\geq$ 15 bei mindestens 2 aufeinanderfolgenden Visiten**

Bis zum Ende der Studie am 11.08.2023 trat laut prüfärztlicher Einschätzung sowie gemäß CERC-Bewertung in keinem der beiden Studienarme ein MELD-Score  $\geq$  15 bei mindestens 2 aufeinanderfolgenden Visiten auf.

Die Erhebung des Endpunkts wird als valide eingeschätzt.

### **4.5 Lebensqualität**

Für den Endpunkt „PBC-40“ liegt für keine der Domänen ein statistisch signifikanter Unterschied zwischen Seladelpar und Placebo zu Monat 12 vor.

Der Endpunkt „PBC-40“ wird im vorliegenden Anwendungsgebiet als valide eingeschätzt.

Aufgrund der geringen Rücklaufquoten zu Monat 12 (Seladelpar: ca. 73 %; Placebo: ca. 79 %) wird das Verzerrungspotenzial als „hoch“ eingestuft.

### **4.6 Sicherheit**

Alle UE wurden ab dem Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis zum Ende der Studienteilnahme (letzte Studienvisite) kontinuierlich erhoben. Die mediane Behandlungsdauer war mit ca. 52 Wochen in beiden Studienarmen vergleichbar. Die mediane Beobachtungsdauer für die Sicherheitsüberwachung lag im Seladelpar-Arm bei 52,4 Wochen (min: 4,4; max: 58,3) und im Placebo-Arm bei 52,1 Wochen (min: 5,6; max: 58,1).

Das relative Risiko (RR) wurde post-hoc berechnet. In Modul 4 wird angegeben, dass sämtliche statistischen Tests – sofern nicht anders angegeben – nach den Baseline-Werten verschiedener Merkmale, einschließlich der Randomisierungsfaktoren, stratifiziert durchgeführt wurden. Es ist unklar, ob tatsächlich alle in Modul 4 aufgeführten Merkmale bei der Stratifikation berücksichtigt wurden, bzw. ob und anhand welcher Merkmale das RR stratifiziert berechnet wurde.

Insbesondere (aber nicht ausschließlich) in Bezug auf die SOC „Erkrankungen der Haut und des Unterhautgewebes“ ist gemäß der aufgetretenen PT davon auszugehen, dass Symptome der Grunderkrankung umfasst sein könnten, da bei knapp 5 % im Seladelpar-Arm sowie knapp 15 % im Placebo-Arm ein Pruritus als UE erfasst wurde.

Für die Gesamtraten der UE liegt kein statistisch signifikanter Unterschied zwischen Seladelpar und Placebo vor. Schwere und schwerwiegende UE traten in keinem der beiden Studienarme auf.

Auf Ebene der Preferred Terms (PT) liegt für den PT „Pruritus“ ein statistisch signifikanter Unterschied zugunsten von Seladelpar gegenüber Placebo vor (RR [95%-KI]: 0,30 [0,12; 0,80],  $p = 0,02$ ). Auch in Bezug auf die AESI liegt für die Pruritus-assoziierten Ereignisse unabhängig vom Schweregrad ein statistisch signifikanter Unterschied zugunsten von Seladelpar gegenüber Placebo vor (RR [95%-KI]: 0,36 [0,14; 0,89],  $p = 0,03$ ). Da Pruritus ein Symptom der Grunderkrankung ist, handelt es sich hierbei vermutlich um Ergebnisse zur Wirksamkeit und nicht um Sicherheitsaspekte. Der Vorteil von Seladelpar gegenüber Placebo in Bezug auf den PT „Pruritus“ und den AESI „Pruritus-assoziierte Ereignisse“ ist somit nicht abschließend bewertbar.

Bei allen weiteren UE nach SOC und PT sowie bei den weiteren AESI zeigten sich keine statistisch signifikanten Unterschiede zwischen Seladelpar und Placebo.

Die Erhebung des Endpunktes wird als weitgehend valide angesehen. Der pU legt in Modul 4 für die Gesamtraten der UE zusätzliche Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Für die UE nach SOC und PT liegen keine entsprechenden Auswertungen vor. Da nicht ersichtlich ist, welche erkrankungsbezogenen Ereignisse dabei nicht berücksichtigt wurden, werden diese Analysen nicht für die Nutzenbewertung herangezogen. Darüber hinaus ist bei der Interpretation der Ergebnisse zu berücksichtigen, dass die berichteten UE möglicherweise auch eine Progression der Grunderkrankung umfassen und auch dadurch teilweise Wirksamkeitsaspekte abgebildet sein können.

Für Teilnehmende, die in die Langzeitnachbeobachtungsstudie (LTE) „ASSURE“ wechselten, entfiel die reguläre Sicherheitsnachbeobachtung von  $14 \pm 3$  Tagen nach Monat 12/EOT. Nach Präspezifikation haben rund 83 % in beiden Studienarmen keine Sicherheitsnachbeobachtung über die Behandlungsphase hinaus wahrgenommen, da sie in die LTE übergegangen sind. Da der Anteil der in die ASSURE-Studie gewechselten Teilnehmenden zwischen den Studienarmen vergleichbar war, wird nicht von einer Verzerrung ausgegangen. Zudem ist unklar, ob adjustierte Effektschätzer vorliegen. Aufgrund der vorliegenden Ergebnisse werden jedoch keine Auswirkungen auf das Verzerrungspotential erwartet. Das Verzerrungspotenzial wird als „niedrig“ eingestuft.

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

Die Angaben der Fachinformation sind zu berücksichtigen.

## 6 Zusammenfassung der Nutzenbewertung

Seladelpar ist zugelassen für die Behandlung der primär biliären Cholangitis (PBC) in Kombination mit Ursodeoxycholsäure (UDCA) bei Erwachsenen, die nicht ausreichend auf UDCA allein ansprechen, oder als Monotherapie bei Patientinnen und Patienten, die UDCA nicht vertragen. Die Nutzenbewertung von Seladelpar basiert auf der zulassungsbegründenden Studie RESPONSE (CB8025-32048). Es handelt sich bei der Studie RESPONSE um eine multizentrische, doppelblinde, randomisierte, Placebo-kontrollierte, Phase-III-Parallelgruppenstudie zur Untersuchung der Wirksamkeit, Verträglichkeit und Pharmakokinetik von Seladelpar bei Erwachsenen mit PBC. Während Teilnehmende mit UDCA-Unverträglichkeit Seladelpar als Monotherapie erhielten, wurde Seladelpar bei Personen, die zu Baseline mit UDCA behandelt wurden, in Kombination mit UDCA verabreicht.

Die Ergebnisse der in der Nutzenbewertung berücksichtigten Endpunkte der Studie werden in der folgenden Tabelle unter Berücksichtigung des Verzerrungspotentials zusammengefasst. Es wird dargestellt, ob unter Seladelpar ein statistisch signifikant positiver Effekt bei hohem oder unklarem ( $\uparrow$ ) oder niedrigem ( $\uparrow\uparrow$ ) Verzerrungspotential bzw. ein statistisch signifikant negativer Effekt bei hohem oder unklarem ( $\downarrow$ ) oder niedrigem ( $\downarrow\downarrow$ ) Verzerrungspotential bzw. kein Unterschied ( $\leftrightarrow$ ) gezeigt werden konnte.

Weitergehende, über die Angaben in der nachfolgenden Zusammenfassenden Tabelle 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 20: Zusammenfassende Darstellung der Ergebnisse der Studie RESPONSE (ITT-Population<sup>1)</sup>)*

Studie RESPONSE Darstellung der Ergebnisse	Seladelpar N = 128			Placebo N = 65			Seladelpar vs. Placebo	Effekt
<b>Gesamtmortalität</b>								
<b>Mortalität</b>	N <sup>2)</sup>	Personen mit Ereignis, n (%)		N <sup>2)</sup>	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert	
Todesfälle	128	0		65	0		-	$\leftrightarrow$
<b>Morbidität</b>								
<b>Pruritus NRS<sup>3)</sup> zu Monat 6</b>	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 6 zu Baseline LS Mean (SE) <sup>4)</sup>	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 6 zu Baseline LS Mean (SE) <sup>4)</sup>	LS-Mean-Differenz [95%-KI]; p-Wert <sup>4)</sup>	
Pruritus NRS – stetige Analyse	127 <sup>5)6)</sup>	3,03 (2,81)	-1,33 (0,14)	64 <sup>5)6)</sup>	3,02 (2,96)	-0,42 (0,19)	-0,90 [-1,35; -0,45]; 0,0001 <sup>7)</sup>	$\uparrow$
<b>Pruritus NRS<sup>3)</sup> zu Monat 6</b>	N <sup>2)</sup>	Personen mit Ereignis zu Monat 6, n (%)		N <sup>2)</sup>	Personen mit Ereignis zu Monat 6, n (%)		RR [95%-KI]; p-Wert	
Pruritus NRS – Responderanalyse	128 <sup>8)</sup>	28 (21,9) <sup>9)10)</sup>		65 <sup>8)</sup>	10 (15,4) <sup>9)10)</sup>		k. A.	n. b.

Studie RESPONSE Darstellung der Ergebnisse	Seladelpar N = 128			Placebo N = 65			Seladelpar vs. Placebo	Effekt
	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 12 zu Baseline LS Mean (SE) <sup>13)</sup>	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 12 zu Baseline LS Mean (SE) <sup>13)</sup>		
5-D Itch <sup>11)12)</sup> zu Monat 12	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 12 zu Baseline LS Mean (SE) <sup>13)</sup>	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 12 zu Baseline LS Mean (SE) <sup>13)</sup>	LS-Mean-Differenz [95%-KI]; p-Wert <sup>13)</sup>	
5-D Itch Gesamtscore	128 <sup>5)14)</sup>	11,6 (4,85)	-2,26 (0,33)	64 <sup>5)14)</sup>	11,2 (4,65)	0,05 (0,46)	-2,31 [-3,40; -1,23]; < 0,0001 <sup>15)</sup>	↑
5-D Itch – Grad	128 <sup>5)14)</sup>	2,26 (1,04)	-0,42 (0,07)	64 <sup>5)14)</sup>	2,24 (0,99)	-0,05 (0,10)	-0,38 [-0,61; -0,15]; 0,002 <sup>16)</sup>	↑
5-D Itch – Dauer	128 <sup>5)14)</sup>	1,59 (1,11)	-0,32 (0,07)	64 <sup>5)14)</sup>	1,56 (1,08)	0,08 (0,10)	-0,40 [-0,63; -0,16]; 0,001 <sup>17)</sup>	↑
5-D Itch – Beeinträchtigung	128 <sup>5)14)</sup>	2,27 (1,37)	-0,55 (0,09)	64 <sup>5)14)</sup>	2,22 (1,24)	0,05 (0,13)	-0,60 [-0,91; -0,29]; 0,0002 <sup>18)</sup>	↑
5-D Itch – Verlauf	128 <sup>5) 14)</sup>	3,35 (1,09)	-0,64 (0,13)	64 <sup>5)14)</sup>	3,13 (1,16)	-0,17 (0,18)	-0,47 [-0,90; -0,04]; 0,03 <sup>19)</sup>	↑
5-D Itch – Lokalisation	128 <sup>5)14)</sup>	2,16 (1,29)	-0,25 (0,09)	64 <sup>5)14)</sup>	2,08 (1,20)	0,14 (0,12)	-0,38 [-0,67; -0,10]; 0,009 <sup>20)</sup>	↑
<b>Klinische PBC- Ereignisse</b>	N <sup>2)</sup>	Personen mit Ereignis, n (%)		N <sup>2)</sup>	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert	
MELD-Score ≥ 15 bei mind. 2 aufeinander- folgenden Visiten <sup>21)</sup>	128	0		65	0		-	↔
<b>Lebensqualität</b>								
<b>PBC-40<sup>22)</sup></b> zu Monat 12	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 12 zu Baseline LS Mean (SE) <sup>23)</sup>	N <sup>2)</sup>	Baseline MW (SD)	Änderung Monat 12 zu Baseline LS Mean (SE) <sup>23)</sup>	LS-Mean-Differenz [95%-KI]; p-Wert <sup>23)</sup>	
Juckreiz	126 <sup>5)24)</sup>	5,14 (3,85)	-1,31 (0,26)	63 <sup>5)24)</sup>	5,60 (3,97)	-0,48 (0,36)	-0,83 (-1,68; 0,02); 0,054	↔
Generelle Symptome	126 <sup>5)24)</sup>	15,1 (4,56)	-0,10 (0,30)	63 <sup>5)24)</sup>	15,7 (5,46)	-0,19 (0,41)	0,09 (-0,87; 1,06); 0,85	↔
Fatigue	126 <sup>5)24)</sup>	27,6 (10,0)	-1,97 (0,67)	63 <sup>5)24)</sup>	27,4 (10,6)	-1,50 (0,92)	-0,47 (-2,66; 1,71); 0,67	↔
Kognitive Funktion	126 <sup>5)24)</sup>	13,2 (5,51)	-0,55 (0,33)	63 <sup>5)24)</sup>	12,8 (4,88)	-0,42 (0,44)	-0,13 (-1,20; 0,94); 0,81	↔
Emotionale Funktion	126 <sup>5)24)</sup>	7,79 (2,89)	-0,70 (0,20)	63 <sup>5)24)</sup>	7,72 (3,15)	-0,95 (0,27)	0,25 (-0,40; 0,90); 0,45	↔
Soziale Funktion	126 <sup>5)24)</sup>	22,9 (8,44)	-0,67 (0,47)	63 <sup>5)24)</sup>	22,2 (8,32)	-1,72 (0,64)	1,05 (-0,47; 2,56); 0,17	↔

Studie RESPONSE Darstellung der Ergebnisse	Seladelpar N = 128		Placebo N = 65		Seladelpar vs. Placebo	Effekt
<b>Sicherheit<sup>25)</sup></b>						
<b>Unerwünschte Ereignisse</b>	<i>N</i> <sup>2)</sup>	<i>Personen mit Ereignis, n (%)</i>	<i>N</i> <sup>2)</sup>	<i>Personen mit Ereignis, n (%)</i>	<i>RR [95%-KI]; p- Wert<sup>26)</sup></i>	
Schwere UE <sup>27)</sup>	128	14 (10,9)	65	5 (7,7)	1,42 [0,54; 3,78]; 0,48	↔
SUE	128	9 (7,0)	65	4 (6,2)	1,14 [0,37; 3,57]; 0,82	↔
UE, das zum Abbruch d. Studienmedikation führte	128	4 (3,1)	65	3 (4,6)	0,68 [0,16; 2,94]; 0,60	↔

<sup>1)</sup> Die ITT-Population entspricht der Sicherheitspopulation (siehe Tabelle 9). Definition, siehe Kapitel 2.4.

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

<sup>3)</sup> Intensität des schlimmsten Juckreizes innerhalb der letzten 24 Stunden auf einer 11-stufigen Skala von 0 (= kein Juckreiz) bis 10 (= schlimmster vorstellbarer Juckreiz) (Selbstbericht).

<sup>4)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-NRS, Randomisierungsstratum (ALP-Wert < 350 U/L versus ALP-Wert ≥ 350 U/L), Behandlungsgruppe, Woche sowie die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline-Wert wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als zusätzlicher Faktor in das Modell aufgenommen.

<sup>5)</sup> Anzahl an Personen, die sowohl einen Baseline-Wert als auch mindestens einen Wert nach Baseline aufwiesen und somit in das MMRM eingeschlossen wurden.

<sup>6)</sup> Verfügbare Daten zu Baseline entsprechen der ITT. Verfügbare Daten zu Monat 6 im Seladelpar-Arm: n = 121 (94,5 %); verfügbare Daten zu Monat 6 im Placebo-Arm: n = 61 (93,8 %).

<sup>7)</sup> Auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz), erscheint das Vorliegen eines klinisch relevanten Effekts plausibel (SMD [95%-KI] = -0,58 [-0,89; -0,28]).

<sup>8)</sup> Im Seladelpar-Arm hatten 98 Personen (76,6 %) und im Placebo-Arm 50 Personen (76,9 %) einen Pruritus-NRS-Wert > 0 zu Baseline.

<sup>9)</sup> Im vorliegenden Fall wurden die Prozentangaben (Eigenberechnung) auf die für die Nutzenbewertung relevante ITT-Population bezogen. Diese umfasst auch Personen mit einem Pruritus-NRS-Wert von 0 zu Baseline, welche in einer Responderanalyse als Non-Responder gewertet werden sollten, in der Analyse des pU jedoch nicht berücksichtigt wurden.

<sup>10)</sup> Den Studienunterlagen konnte nicht entnommen werden, wie viele der Personen ohne Ereignis tatsächlich Non-Responder waren und wie viele zu Monat 6 fehlende Werte aufwiesen oder die Studie vorzeitig abgebrochen haben. Über den gesamten Studienzeitraum hinweg brachen 11 Personen im Seladelpar-Arm und 8 Personen im Placebo-Arm die Studie vorzeitig ab. Allerdings ist unklar, wie viele dieser Abbrüche bereits innerhalb der ersten 6 Monate stattgefunden haben.

<sup>11)</sup> Die zu erreichende Gesamtpunktzahl liegt zwischen 5 und 25 Punkten. Jede Domäne kann einen Wert von 1 bis 5 Punkten erreichen. Höhere Punkte deuten auf einen stärkeren Juckreiz hin.

<sup>12)</sup> Die mediane Beobachtungsdauer für die 5-D Itch lag in beiden Studienarmen bei 44 Wochen (für Seladelpar min: 4,1; max: 46; für Placebo min: 0,1; max: 45,1). Bei der Interpretation der Ergebnisse zu Monat 12 ist somit zu berücksichtigen, dass keine tatsächlichen Werte zu Woche 52 vorlagen.

<sup>13)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-Werten, Randomisierungsstratum zu Studienbeginn (ALP-Wert < 350 U/L vs. ALP-Wert ≥ 350 U/L und Pruritus NRS < 4 vs. Pruritus NRS ≥ 4), Behandlungsgruppe, Woche und die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline-Wert wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als zusätzlicher Faktor in das Modell aufgenommen.

<sup>14)</sup> Verfügbare Daten zu Baseline entsprechen der ITT. Verfügbare Daten zu Monat 12 im Seladelpar-Arm: n = 96 (75 %); verfügbare Daten zu Monat 12 im Placebo-Arm: n = 46 (70,8 %).

<sup>15)</sup> Auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz), erscheint das Vorliegen eines klinisch relevanten Effekts plausibel (SMD [95%-KI] = -0,62 [-0,93; -0,31]).

- <sup>16)</sup> Es lässt sich auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz) mit einem Wert von -0,47 (95%-KI: [-0,77; -0,16]) nicht ableiten, dass ein klinisch relevanter Effekt vorliegt.
- <sup>17)</sup> Es lässt sich auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz) mit einem Wert von -0,49 (95%-KI: [-0,79; -0,19]) nicht ableiten, dass ein klinisch relevanter Effekt vorliegt.
- <sup>18)</sup> Auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz), erscheint das Vorliegen eines klinisch relevanten Effekts plausibel (SMD [95%-KI] = -0,57 [-0,87; -0,26]).
- <sup>19)</sup> Es lässt sich auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz) nicht ableiten, dass ein klinisch relevanter Effekt vorliegt (SMD [95%-KI] = -0,32 [-0,62; -0,02]).
- <sup>20)</sup> Es lässt sich auf Basis von Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz) nicht ableiten, dass ein klinisch relevanter Effekt vorliegt (SMD [95%-KI] = -0,39 [-0,69; -0,09]).
- <sup>21)</sup> Gemäß CERC-Bewertung; Einschätzungen gemäß prüfärztlicher Einschätzung stimmen überein.
- <sup>22)</sup> Erhebung der gesundheitsbezogenen Lebensqualität bei Teilnehmenden mit PBC, bestehend aus 40 Items in 6 Domänen. Je nach Anzahl der Items in den einzelnen Domänen ergeben sich unterschiedliche Wertebereiche für die Scores der Domänen, die von 3 bis 55 Punkten reichen. Höhere Punktzahlen spiegeln eine Verschlechterung wider, während niedrigere Punktzahlen eine Verbesserung bedeuten.
- <sup>23)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline-Werten, Randomisierungsstratum zu Studienbeginn (ALP-Wert < 350 U/L vs. ALP-Wert ≥ 350 U/L und NRS < 4 vs. NRS ≥ 4), Behandlungsgruppe, Woche und die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline des PBC-40-Gesamtscores wurde untersucht und aufgrund ihres signifikanten Effekts (p-Wert < 0,05) als Term in die Modellanalyse für die Domäne „Soziale Funktion“ aufgenommen.
- <sup>24)</sup> Verfügbare Daten zu Baseline entsprechen der ITT. Verfügbare Daten zu Monat 12 im Seladelpar-Arm: n = 94 (73,4 %); verfügbare Daten zu Monat 12 im Placebo-Arm: n = 51 (78,5 %).
- <sup>25)</sup> Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind im Ergebniskapitel 3.5 Sicherheit dargestellt.
- <sup>26)</sup> Das RR wurde post-hoc berechnet anhand der Mantel-Haenszel-Methode. Die 95%-Konfidenzintervalle und p-Werte wurden unter Verwendung der Normalapproximation (Wald-Test) berechnet. Es ist nicht eindeutig ersichtlich, ob das RR stratifiziert nach den Randomisierungsfaktoren berechnet wurde. Ein RR < 1 bedeutet einen Vorteil für die Behandlung mit Seladelpar.
- <sup>27)</sup> Die Schweregradeinteilung erfolgte gemäß CTCAE (Version 5.0).

Abkürzungen: ALP: Alkalische Phosphatase; CTCAE: Common Terminology Criteria for Adverse Events; ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MELD: Model for End-Stage Liver Disease; MMRM: Gemischtes Modell mit Messwiederholungen (Mixed Effect Model for Repeated Measurements); MW: Mittelwert; n. a.: nicht anwendbar; n. b.: nicht bewertbar; N: Anzahl; NRS: Numerical Rating Scale; PBC: Primär biliäre Cholangitis; RR: Relatives Risiko; SD: Standardabweichung; SE: Standardfehler; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## Referenzen

1. **Berg T, Aehling N, Bruns T, Welker M, Weismüller T, Trebicka J, et al.** S2k-Leitlinie Lebertransplantation der Deutschen Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS) und der Deutschen Gesellschaft für Allgemein- und Viszeralchirurgie (DGAV); Langfassung [online]. AWMF-Registernummer: 021–029. Berlin (GER): Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF); 2023. [Zugriff: 26.05.2025]. URL: [https://register.awmf.org/assets/guidelines/021-029l\\_S2k\\_Lebertransplantation\\_2024-07.pdf](https://register.awmf.org/assets/guidelines/021-029l_S2k_Lebertransplantation_2024-07.pdf).
2. **CymaBay Therapeutics.** ASSURE: an open label long-term study to evaluate the safety and tolerability of seladelpar in subjects with primary biliary cholangitis (PBC); CB8025-31731-RE version 6.1-EU; protocol [unveröffentlicht]. 14.11.2024.
3. **CymaBay Therapeutics.** RESPONSE: a placebo-controlled, randomized, phase 3 study to evaluate the efficacy and safety of seladelpar in patients with primary biliary cholangitis (PBC) and an inadequate response to or an intolerance to ursodeoxycholic acid (UDCA); CB8025-32048 version 4.1.DE; protocol [unveröffentlicht]. 14.02.2022.
4. **CymaBay Therapeutics.** RESPONSE: a placebo-controlled, randomized, phase 3 study to evaluate the efficacy and safety of seladelpar in patients with primary biliary cholangitis (PBC) and an inadequate response to or an intolerance to ursodeoxycholic acid (UDCA); CB8025-32048; clinical study report [unveröffentlicht]. 30.11.2023.
5. **CymaBay Therapeutics.** RESPONSE: a placebo-controlled, randomized, phase 3 study to evaluate the efficacy and safety of seladelpar in patients with primary biliary cholangitis (PBC) and an inadequate response to or an intolerance to ursodeoxycholic acid (UDCA); CB8025-32048; statistical analysis plan version 1.0 [unveröffentlicht]. 28.08.2023.
6. **CymaBay Therapeutics.** RESPONSE: a placebo-controlled, randomized, phase 3 study to evaluate the efficacy and safety of seladelpar in patients with primary biliary cholangitis (PBC) and an inadequate response to or an intolerance to ursodeoxycholic acid (UDCA); CB8025-32048; statistical analysis plan version 1.0 supplement [unveröffentlicht]. 29.08.2023.
7. **European Medicines Agency (EMA).** Seladelpar Gilead (seladelpar): European public assessment report EMEA/H/C/004692/0000 [online]. 12.12.2024. Amsterdam (NED): EMA. [Zugriff: 02.04.2025]. URL: [https://www.ema.europa.eu/en/documents/assessment-report/seladelpar-gilead-epar-public-assessment-report\\_en.pdf](https://www.ema.europa.eu/en/documents/assessment-report/seladelpar-gilead-epar-public-assessment-report_en.pdf).
8. **Gilead Sciences.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4A: Anhang 4-G; Seladelpar (Seladelpar Gilead), zur Behandlung der primär biliären Cholangitis (PBC) in Kombination mit UDCA bei Erwachsenen, die nicht ausreichend auf UDCA alleine ansprechen, oder als Monotherapie bei Patienten, die UDCA nicht vertragen [unveröffentlicht]. 14.03.2025.
9. **Gilead Sciences.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4A: Seladelpar (Seladelpar Gilead), zur Behandlung der primär biliären Cholangitis (PBC) in Kombination mit UDCA bei Erwachsenen, die nicht ausreichend auf UDCA alleine ansprechen, oder als Monotherapie bei Patienten, die UDCA nicht vertragen; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 14.03.2025.

10. **Gilead Sciences.** Seladelpar Gilead 10 mg Hartkapseln [online]. 02.2025. Frankfurt/Main. [Zugriff: 02.04.2025]. URL: <https://www.fachinfo.de/>.
11. **Jacoby A, Rannard A, Buck D, Bhala N, Newton JL, James OF, et al.** Development, validation, and evaluation of the PBC-40, a disease specific health related quality of life measure for primary biliary cirrhosis. *Gut* 2005;54(11):1622-9.
12. **Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.** Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-766 e1.
13. **Newcastle University.** PBC-40 [online]. Newcastle upon Tyne (GBR): UK-PBC 2025. [Zugriff: 26.05.2025]. URL: <https://www.uk-pbc.com/resources/tools/pbc-40/>.
14. **Nixon A, Doll H, Kerr C, Burge R, Naegeli AN.** Interpreting change from patient reported outcome (PRO) endpoints: patient global ratings of concept versus patient global ratings of change, a case study among osteoporosis patients. *Health Qual Life Outcomes* 2016;14:25.
15. **Norman GR, Stratford P, Regehr G.** Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50(8):869-79.

## Anhang

### Kombiniertes biochemisches Ansprechen

Die mediane Beobachtungsdauer für ALP und Bilirubin betrug in beiden Studienarmen 52,1 Wochen (Seladelpar min: 5,0; max: 59,0 und Placebo min: 6,3; max: 59,3).

*Tabelle 21: Biochemisches Ansprechen zu Monat 12, Responderanalysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>1)</sup>*

Studie RESPONSE Biochemisches Ansprechen	Seladelpar N = 128	Placebo N = 65	RR <sup>2)</sup> [95%-KI]; p-Wert
<b>Biochemisches Ansprechen (ALP &lt; 1,67 × ULN und TB ≤ ULN und ALP-Senkung ≥ 15 %)<sup>3)</sup></b>			
n (%) Ereignisse, n (%) <sup>4)</sup>	128 (100) 79 (61,7) <sup>5)</sup>	65 (100) 13 (20,0) <sup>5)</sup>	3,09 [1,87; 5,10]; < 0,0001
<b>ALP &lt; 1,67 × ULN</b>			
n (%) Ereignisse, n (%) <sup>4)</sup>	128 (100) 84 (65,6) <sup>5)</sup>	65 (100) 17 (26,2) <sup>5)</sup>	2,51 [1,65; 3,80]; < 0,0001
<b>ALP-Reduktion ≥ 15 %</b>			
n (%) Ereignisse, n (%) <sup>4)</sup>	128 (100) 107 (83,6) <sup>5)</sup>	65 (100) 21 (32,3) <sup>5)</sup>	2,61 [1,81; 3,75]; < 0,0001
<b>TB ≤ ULN</b>			
n (%) Ereignisse, n (%) <sup>4)</sup>	128 (100) 104 (81,3) <sup>5)</sup>	65 (100) 50 (76,9) <sup>5)</sup>	1,06 [0,91; 1,24]; 0,45

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> Das RR wurde post-hoc berechnet anhand der Mantel-Haenszel-Methode. Die 95%-Konfidenzintervalle und p-Werte wurden unter Verwendung der Normalapproximation (Wald-Test) berechnet. Es erfolgte eine Adjustierung für die Randomisierungsstrata. Ein RR < 1 bedeutet einen Vorteil für die Behandlung mit Seladelpar.

<sup>3)</sup> Eine Person wurde als Responder eingestuft, wenn alle 3 der folgenden Bedingungen erfüllt waren: ALP < 1,67 × ULN; ALP-Abnahme gegenüber dem Ausgangswert um ≥ 15 %; Gesamt-Bilirubin ≤ 1,0 × ULN.

<sup>4)</sup> Personen, für die keine Bewertung vorliegt, die die Behandlung vor Monat 12 abgebrochen haben oder aus anderen Gründen fehlende Werte aufweisen, werden als Non-Responder gewertet.

<sup>5)</sup> Insgesamt wiesen im Seladelpar-Arm 14 Personen und im Placebo-Arm 8 Personen fehlende Werte zu Monat 12 auf und wurden somit als Non-Responder imputiert.

Abkürzungen: ALP: Alkalische Phosphatase; ITT: Intention to treat; RR: Relatives Risiko; TB: Gesamtbilirubin; ULN: Oberer Normwert.

## Pruritus

### PGI-S

Die mediane Beobachtungsdauer für den PGI-S betrug 51,9 Wochen im Seladelpar-Arm (min: 0,1; max: 54,3) und 52 Wochen im Placebo-Arm (min: 0,1; max: 55,4).

Tabelle 22: PGI-S zu Monat 12; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>1)</sup>

Studie RESPONSE PGI-S <sup>2)</sup>	Seladelpar N = 128 n (%)	Placebo N = 65 n (%)
<i>Baseline</i> <sup>3)</sup>	128 (100)	64 (98,5)
kein Juckreiz	43 (33,6)	19 (29,2)
mild	35 (27,3)	21 (32,3)
moderat	40 (31,3)	19 (29,2)
schwer	10 (7,8)	5 (7,7)
<i>Monat 12</i>	94 (73,4)	51 (78,5)
kein Juckreiz	46 (35,9)	24 (36,9)
mild	32 (25,0)	15 (23,1)
moderat	14 (10,9)	9 (13,8)
schwer	2 (1,56)	3 (4,61)
<i>Anteil Personen mit Verbesserung zu Monat 12</i>	k. A.	k. A.

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> Patienteneinschätzung des Ausmaßes des Juckreizes.

<sup>3)</sup> Baseline war definiert als arithmetisches Mittel aller verfügbaren Messungen bei Screening, Run-in, Tag 1 und ggf. ungeplanten Erhebungen vor oder an Tag 1.

Abkürzungen: ITT: Intention to Treat; PGI-S: Patient Global Impression of Severity.

### PGI-C

Die mediane Beobachtungsdauer für den PGI-C betrug 51,9 Wochen im Seladelpar-Arm (min: 0,1; max: 54,3) und 52 Wochen im Placebo-Arm (min: 0,1; max: 55,4).

Tabelle 23: Veränderung des PGI-C zu Monat 12 gegenüber Monat 1, MMRM-Analysen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>1)</sup>

Studie RESPONSE PGI-C <sup>2)</sup> zu Monat 12	Seladelpar N = 128	Placebo N = 65
<i>Monat 1</i> <sup>3)</sup>		
n (%)	116 (90,6)	61 (93,8)
MW (SD)	3,38 (1,08)	3,44 (1,27)
<i>Monat 12</i>		
n (%)	94 (73,4)	51 (78,5)
MW (SD)	2,64 (1,42)	3,12 (1,18)
<i>Veränderung von Monat 1 zu Monat 12</i>		
n (%) <sup>4)</sup>	124 (96,9) <sup>6)</sup>	63 (96,9) <sup>6)</sup>
LS Mean (SE) <sup>5)</sup>	2,60 (0,14)	3,01 (0,18)
LS Mean Difference [95%-KI]; p-Wert <sup>6)</sup>		-0,41 [-0,85, 0,03]; 0,07

- <sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.
- <sup>2)</sup> Patienteneinschätzung des Ausmaßes der Veränderung des Juckreizes im Vergleich zum Studienbeginn.
- <sup>3)</sup> Aufgrund der Formulierung des Items ist keine Baseline-Erhebung möglich bzw. vorgesehen.
- <sup>4)</sup> Anzahl an Personen, die sowohl einen Baseline-Wert als auch mindestens einen Wert nach Baseline aufwiesen und somit in das MMRM eingeschlossen wurden.
- <sup>5)</sup> MMRM-Modell mit unstrukturierter Kovarianzstruktur sowie Baseline (in diesem Fall Monat 1), Randomisierungsstratum (ALP-Wert < 350 U/L versus ALP-Wert ≥ 350 U/L), Behandlungsgruppe, Woche sowie die Interaktion zwischen Behandlung und Woche als Faktoren. Für die Freiheitsgrade im Nenner erfolgte eine Kenward-Roger-Korrektur. Die Interaktion zwischen Behandlung und Baseline-Wert (hier Monat 1) wurde untersucht und als zusätzlicher Faktor aufgenommen, falls signifikant (p-Wert < 0,05). Es ist nicht ersichtlich, ob dies im vorliegenden Fall zutrifft.
- <sup>6)</sup> Prozentangaben aus eigener Berechnung.

Abkürzungen: ALP: Alkalische Phosphatase; ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MMRM: Gemischtes Modell mit Messwiederholungen; MW: Mittelwert; NRS: Numerical Rating Scale; SD: Standardabweichung; SE: Standardfehler.

## Klinische PBC-Ereignisse

### Lebertransplantation

Bis zum Ende der Studie am 11.08.2023 trat laut prüfärztlicher Einschätzung sowie gemäß CERC-Bewertung in keinem der beiden Studienarme eine Lebertransplantation auf.

### Behandlungsbedürftiger Aszites

Bis zum Ende der Studie am 11.08.2023 trat laut prüfärztlicher Einschätzung sowie gemäß CERC-Bewertung in keinem der beiden Studienarme ein behandlungsbedürftiger Aszites auf. Bei einer Person wurde ein möglicher behandlungsbedürftiger Aszites durch das CERC bewertet, jedoch nicht als „positiv“ kategorisiert.

### Hospitalisierung

Tabelle 24: Hospitalisierungen; Studie RESPONSE, ITT-Population; Datenschnitt 11.08.2023<sup>1)</sup>

Studie RESPONSE Hospitalisierung <sup>2)</sup>	Seladelpar N = 128 n (%)	Placebo N = 65 n (%)	RR [95%-KI]; p-Wert
Prüfärztliche Einschätzung	1 (0,8)	0	k. A.
Durch das CERC beurteilt <sup>3)</sup>	6 (4,7)	3 (4,6)	1,06 [0,27; 4,17]; 0,9
Durch das CERC als „positiv“ eingestuft <sup>4)</sup>	1 (0,8)	0	k. A.

<sup>1)</sup> Es liegen keine eindeutigen Angaben zum Zeitpunkt des Datenschnitts in den Studienunterlagen vor. Die vorliegende Angabe entstammt Modul 4.

<sup>2)</sup> Erfasst wurden Hospitalisierungen aufgrund eines neuen Auftretens oder eines Wiederauftretens von: Varizenblutung, Hepatische Enzephalopathie oder spontane bakterielle Peritonitis.

<sup>3)</sup> In Modul 4 werden die Ergebnisse, für die vom CERC bewerteten Ereignisse präsentiert. Es bleibt jedoch unklar, ob hierbei sämtliche Fälle erfasst sind, die potenziell als PBC-Ereignis in Frage kamen und deshalb dem CERC zur Beurteilung vorgelegt wurden, oder was genau unter „beurteilt“ in diesem Zusammenhang zu verstehen ist.

<sup>4)</sup> Die CERC-Mitglieder bewerteten im Rahmen einer verblindeten, unabhängigen Abstimmung, ob die Definitionen für klinische PBC-Ereignisse erfüllt waren. Es erfolgte eine Einordnung in die Kategorien „positiv“ (also Ereignis liegt vor), „negativ“ (Ereignis liegt nicht vor) oder „unbestimmt“.

Abkürzungen: CERC: Clinical Event Review Committee; ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; PBC: Primär biliäre Cholangitis; RR: Relatives Risiko.