

# 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: Avapritinib**

**Neues Anwendungsgebiet: Indolente systemische Mastozytose**

Datum der Veröffentlichung: 2. April 2024

## Inhaltsverzeichnis

Tabellenverzeichnis .....	3
Abkürzungsverzeichnis .....	5
Hintergrund.....	6
1 Fragestellung.....	7
2 Auswahl und Methodik der relevanten Studien .....	8
2.1 Studienbasis für die Nutzenbewertung .....	8
2.2 Beschreibung der eingeschlossenen Studie .....	8
2.3 Endpunkte .....	15
2.3.1 Mortalität.....	16
2.3.2 Morbidität.....	16
2.3.3 Lebensqualität .....	23
2.3.4 Sicherheit .....	25
2.3.5 Erhebungszeitpunkte .....	28
2.4 Statistische Methoden.....	28
2.5 Verzerrungspotential auf Studien- und Endpunktebene .....	31
3 Ergebnisse der eingeschlossenen Studie .....	33
3.1 Studiencharakteristika und Studienmedikation .....	33
3.2 Mortalität .....	36
3.3 Morbidität .....	36
3.3.1 Symptome mittels „ISM-SAF“ und „PGIS“ .....	36
3.3.2 Allgemeiner Gesundheitszustand mittels „EQ-5D-VAS“ .....	38
3.3.3 Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin.....	39
3.4 Lebensqualität.....	39
3.5 Sicherheit .....	41
4 Diskussion der Methodik und Ergebnisse.....	44
4.1 Zulassungsstatus und Zulassungspopulation von Avapritinib .....	44
4.2 Design und Methodik der Studie.....	44
4.3 Mortalität .....	46
4.4 Morbidität .....	46
4.5 Lebensqualität.....	47
4.6 Sicherheit .....	47
5 Anforderung an eine qualitätsgesicherte Anwendung .....	49
6 Zusammenfassung der Nutzenbewertung .....	50
Referenzen .....	54
Anhang .....	56

## Tabellenverzeichnis

Tabelle 1: Übersicht über die Studienbasis .....	8
Tabelle 2: Charakterisierung der Studie PIONEER .....	8
Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie PIONEER.....	13
Tabelle 4: Charakterisierung der Intervention Avapritinib.....	14
Tabelle 5: Zusammenfassung der Endpunktbewertung der Studie PIONEER .....	15
Tabelle 6: Items und Domänen des ISM-SAF .....	17
Tabelle 7: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der 24-wöchigen RCT-Phase (Teil 2) der Studie PIONEER.....	28
Tabelle 8: Verzerrungspotential der 24-wöchigen RCT-Phase (Teil 2) der Studie PIONEER.....	31
Tabelle 9: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie PIONEER .....	32
Tabelle 10: Allgemeine Angaben; Studie PIONEER .....	33
Tabelle 11: Charakterisierung der Studienpopulation; Studie PIONEER, ITT-Population.....	34
Tabelle 12: Symptome mittels ISM-SAF; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022).....	36
Tabelle 13: Symptome mittels PGIS; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022) .....	38
Tabelle 14: Allgemeiner Gesundheitszustand mittels EQ-5D-VAS; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022).....	38
Tabelle 15: Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022).....	39
Tabelle 16: Gesundheitsbezogene Lebensqualität mittels SF-12; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022).....	39
Tabelle 17: Krankheitsspezifische Lebensqualität mittels MC-QoL; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022).....	40
Tabelle 18: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022).....	41
Tabelle 19: UE mit Inzidenz $\geq 10\%$ ; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022).....	41
Tabelle 20: Schwere UE (CTCAE-Grad 3 oder höher), die bei $\geq 5\%$ der Patientinnen und Patienten in mindestens einem Studienarm aufgetreten sind, nach Systemorgan- klasse; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022) .....	42
Tabelle 21: UESI nach Systemorganklasse und Preferred Term; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022).....	43
Tabelle 22: Ödeme als UESI nach Customized MedDRA Queries; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022).....	43
Tabelle 23: Zusammenfassende Darstellung der Ergebnisse der Studie PIONEER.....	51

Tabelle 24: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite des ISM-SAF zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	57
Tabelle 25: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite des PGIS zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	58
Tabelle 26: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite der EQ-5D-VAS zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	58
Tabelle 27: Mittlere Veränderung des Gesundheitszustands mittels PGIC; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022) .....	58
Tabelle 28: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite des PGIC zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	59
Tabelle 29: Gesundheitsbezogene Lebensqualität mittels SF-12; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	59
Tabelle 30: Lebensqualität mittels MC-QoL; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	60
Tabelle 31: Gesamtraten der UE bis Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022) .....	61

## Abkürzungsverzeichnis

AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ANCOVA	Kovarianzanalyse (Analysis of Covariance)
BSC	Best Supportive Care
CTCAE	Common Terminology Criteria for Adverse Events
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EQ-5D-VAS	Visuelle Analogskala des European Quality of Life 5-Dimension
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GSS	Gastrointestinaler Symptomscore
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISM	Indolente systemische Mastozytose
ISM-SAF	Indolent Systemic Mastocytosis Symptom Assessment Form
ITT	Intention-to-Treat
MC-QoL	Mastocytosis Quality of Life Questionnaire
MCS	Mental Component Summary
MedDRA	Medical Dictionary for Regulatory Activities
MWD	Mittelwertdifferenz
N	Anzahl
NSS	Neurokognitives Symptomcluster
PCS	Physical Component Summary
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PP	Per Protokoll
PRO	Patientenberichteter Endpunkt (Patient-Reported Outcome)
pU	pharmazeutischer Unternehmer
RCT	Randomisierte kontrollierte Studie (Randomized Controlled Trial)
SAP	Statistischer Analyseplan
SF-12	Short Form 12 Health Survey
SGB	Sozialgesetzbuch
SSS	Haut-Symptomscore (Skin Symptom Score)
SUE	Schwerwiegendes unerwünschtes Ereignis
TSS	Gesamt-Symptomscore (Total Symptom Score)
UE	Unerwünschtes Ereignis
UESI	Unerwünschtes Ereignis von besonderem Interesse
VerfO	Verfahrensordnung des G-BA

## Hintergrund

Avapritinib 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 Avapritinib 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 Avapritinib in seiner Sitzung am 26. März 2024 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 21. Dezember 2023 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pU in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 2. April 2024 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

## 1 Fragestellung

Avapritinib (AYVAKYT®) 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 [4]:

Behandlung erwachsener Patienten mit indolenter systemischer Mastozytose (ISM) mit mittelschweren bis schweren Symptomen, bei denen mit einer symptomatischen Behandlung keine ausreichende Kontrolle erzielt werden kann.

Gemäß Fachinformation beträgt die zugelassene Dosierung von Avapritinib 25 mg, die einmal täglich oral eingenommen werden soll. Die Behandlung von ISM sollte fortgesetzt werden, bis die Erkrankung fortschreitet oder eine inakzeptable Toxizität auftritt. [4]

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studiename (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>				
PIONEER (BLU-285-2203) <sup>1)</sup>	Ja	Ja	Ja	--

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

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

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

#### Zur Nutzenbewertung für Avapritinib herangezogene Unterlagen

- Dossier des pU zu Avapritinib [5]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [6]
- Studienbericht inkl. Studienprotokoll, statistischem Analyseplan (SAP) und statistischen Outputs der Studie PIONEER (BLU-285-2203) [1,2,3]
- Fachinformation zu Avapritinib [4]
- Gotlib et al., 2023 [7]

### 2.2 Beschreibung der eingeschlossenen Studie

Die Nutzenbewertung von Avapritinib basiert im vorliegenden Anwendungsgebiet auf der Zulassungsstudie PIONEER (BLU-285-2203). Die Studie und die Intervention werden in den Tabellen 2–4 charakterisiert.

Tabelle 2: Charakterisierung der Studie PIONEER

Charakteristikum	Beschreibung
<b>Design und Studienablauf</b>	<p><b>Phase-II-Studie</b></p> <ul style="list-style-type: none"> <li>• Behandlung von Patientinnen und Patienten mit ISM, deren Symptome durch BSC nicht ausreichend kontrolliert werden konnten.</li> <li>• Dreiteilige Studie zur <ul style="list-style-type: none"> <li>○ Bestimmung der Avapritinib-Dosis (Teil 1),</li> <li>○ Untersuchung der Wirksamkeit und Sicherheit von Avapritinib in Kombination mit BSC im Vergleich zu Placebo in Kombination mit BSC (Teil 2) sowie</li> <li>○ Bestimmung der Langzeitwirksamkeit und -sicherheit von Avapritinib (Teil 3)</li> </ul> </li> <li>• Teil 1 und 2 (RCT-Phasen) haben unterschiedliche Patientinnen und Patienten, Teil 3 entspricht der einarmigen Langzeit-Extensionsphase zu Teil 1 und 2.</li> </ul>



Charakteristikum	Beschreibung
	<p><u>Screening-Phase bei Teil 1 und 2</u></p> <ul style="list-style-type: none"> <li>• Dauer insgesamt: bis zu 14 Wochen.</li> <li>• Tägliches Ausfüllen des Fragebogens ISM-SAF durch die Patientinnen und Patienten.</li> </ul> <ol style="list-style-type: none"> <li>1. Optimierung und Stabilisierung der BSC-Therapie (bis zu 4 Wochen).</li> <li>2. Feststellung der Studieneignung bei optimierter BSC-Therapie (mittlerer ISM-SAF TSS <math>\geq 28</math> bei 14-tägiger Erhebung).</li> <li>3. Screening-Assessment: Pathologie, Labor, diagnostische Tests (4 Wochen).</li> <li>4. Bewertung der Ergebnisse und ggfs. Aufnahme der Patientin / des Patienten in die Studie, Bestimmung des ISM-SAF TSS zu Baseline während der letzten 14 Tage vor Behandlungsbeginn<sup>1)</sup> (4 Wochen).</li> </ol> <p><u>Teil 1: Randomisierte doppelblinde Dosisfindungsphase</u></p> <ul style="list-style-type: none"> <li>• 4-armig (25 mg, 50 mg, 100 mg Avapritinib, Placebo).</li> <li>• Behandlungsphase mind. 12 Wochen (3 Zyklen á 28 Tage), bis die empfohlene Phase-II-Dosierung bestimmt wurde<sup>2)</sup>.</li> </ul> <p><u>Teil 2: Randomisiert-kontrollierte doppelblinde Studienphase zur Untersuchung der Wirksamkeit und Sicherheit von Avapritinib + BSC vs. Placebo + BSC</u></p> <ul style="list-style-type: none"> <li>• 2:1-Randomisierung, stratifiziert nach Serumtryptase-Level (<math>&lt; 20</math> ng/ml vs. <math>\geq 20</math> ng/ml); Begrenzung der Teilnehmenden mit Serumtryptase <math>&lt; 20</math> ng/ml auf ca. 20 %.</li> <li>• Behandlungsphase mind. 24 Wochen (6 Zyklen á 28 Tage).</li> </ul> <p><u>Teil 3: Langzeitextensionsphase zur Untersuchung der Langzeitsicherheit und -wirksamkeit</u></p> <ul style="list-style-type: none"> <li>• Einarmig, open label.</li> <li>• Nach Teil 1 oder Teil 2 war ein Übergang in Teil 3 möglich.</li> <li>• Personen des Placebo-Arms konnten in den Interventionsarm wechseln.</li> <li>• Max. 5 Jahre, inklusive Behandlungszeit aus Teil 1 und Teil 2.</li> </ul> <p>Zum Zeitpunkt der Nutzenbewertung sind Teil 1 und Teil 2 der Studie abgeschlossen; Teil 3 der Studie läuft noch.</p> <p>Zur Nutzenbewertung wird die 24-wöchige doppelblinde RCT-Phase (Teil 2) herangezogen, aufgrund des randomisiert-kontrollierten Studiendesigns und der Behandlungsdauer von 24 Wochen.</p>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Personen <math>\geq 18</math> Jahre.</li> <li>• Personen mit bestätigter SM gemäß den WHO-Diagnosekriterien (siehe Anhang).</li> <li>• Mittelschwere bis schwere Symptome, basierend auf einem mittleren ISM-SAF TSS <math>\geq 28</math>, erhoben während des 14-tägigen Screenings zur Studieneignung.</li> <li>• Das Erreichen einer Symptomkontrolle über ein- oder mehrere Symptome war durch mind. 2 der folgenden symptomatischen Therapien nicht möglich: <ul style="list-style-type: none"> <li>○ H1-Blocker</li> <li>○ H2-Blocker</li> <li>○ Protonenpumpenhemmer</li> <li>○ Leukotrien-Inhibitoren</li> <li>○ Cromoglicinsäure</li> <li>○ Kortikosteroide</li> <li>○ Omalizumab</li> </ul> </li> <li>• Stabile symptomatische SM-Therapien der Patientin / des Patienten (z. B. H1- und H2-Blocker): gleiche Dosis, keine neuen Medikamente <math>\geq 14</math> Tage vor Beginn der 14-tägigen ISM-SAF-Bestimmung des TSS zur Testung der Teilnahmeignung.</li> <li>• Bei Personen, die Kortikosteroide erhielten, musste die Dosis <math>\leq 20</math> mg/d Prednison oder Äquivalent und <math>\geq 14</math> Tage vor Beginn der 14-tägigen ISM-SAF-Bestimmung des TSS zur Testung der Teilnahmeignung sein.</li> <li>• ECOG-PS von 0 bis 2.</li> </ul>

Charakteristikum	Beschreibung
	<p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Personen, die mit einer der folgenden WHO-SM-Subklassifikationen diagnostiziert wurden: <ul style="list-style-type: none"> <li>○ Nur kutane Mastozytose (CM), d. h. ohne Dokumentation einer systemischen Beteiligung</li> <li>○ Schwelende systemische Mastozytose (SSM)</li> <li>○ Systemische Mastozytose mit assoziierter hämatologischer Neoplasie (SM-AHN)</li> <li>○ Aggressive systemische Mastozytose (ASM)</li> <li>○ Mastzelleukämie (MCL)</li> <li>○ Mastzellsarkom</li> </ul> </li> <li>• Eine andere myeloproliferative Erkrankung (z. B. myelodysplastisches Syndrom, myeloproliferative Neoplasie)</li> <li>• C-Befunde für Organschäden, die auf SM zurückzuführen sind: <ul style="list-style-type: none"> <li>○ Zytopenie</li> <li>○ Hepatomegalie mit Aszites und eingeschränkter Leberfunktion</li> <li>○ Palpable Splenomegalie mit Hypersplenismus</li> <li>○ Malabsorption mit Hypoalbuminämie und erheblichem Gewichtsverlust</li> <li>○ Skelettveränderungen: große osteolytische Läsionen mit pathologischen Frakturen</li> <li>○ Lebensbedrohliche Organschäden in anderen Organsystemen, die durch Mastzell-Infiltration in Gewebe verursacht werden</li> </ul> </li> <li>• Erfüllen folgender Laborkriterien: <ul style="list-style-type: none"> <li>○ AST oder ALT &gt; 3,0 × ULN</li> <li>○ Gesamtbilirubin &gt; 1,5 × ULN</li> <li>○ Albumin &lt; 1 × LLN</li> <li>○ Geschätzte glomeruläre Filtrationsrate &lt; 30 ml/min/1,73 m<sup>2</sup> oder Kreatinin &gt; 1,5 × ULN</li> <li>○ Absolute Neutrophilenzahl &lt; 1,5 × 10<sup>9</sup>/l</li> <li>○ Hämoglobin &lt; 10 g/dl</li> <li>○ Thrombozytenzahl &lt; 100.000 µl</li> </ul> </li> <li>• Erhalt eines der folgenden Medikamente, Therapien oder Verfahren in den angegebenen Zeiträumen: <ul style="list-style-type: none"> <li>○ Tyrosinkinase-Inhibitoren, einschl., aber nicht beschränkt auf Masitinib und Midostaurin, oder ein Prüfpräparat für &lt; 14 Tage oder 5 Halbwertszeiten des Medikaments.</li> <li>○ Jegliche antineoplastische Arzneimitteltherapie (einschl., aber nicht beschränkt auf Cladribin und Interferon alpha, pegyliertes Interferon oder Antikörpertherapie) &lt; 28 Tage oder 5 Halbwertszeiten des Medikaments.</li> <li>○ Strahlentherapie oder Therapie mit PUVA &lt; 14 Tage vor Beginn der 14-tägigen ISM-SAF-Bestimmung des TSS zur Testung der Teilnahmeeignung.</li> <li>○ Behandlung mit hämatopoetischen Wachstumsfaktoren &lt; 14 Tage vor Beginn der 14-tägigen ISM-SAF-Bestimmung des TSS zur Testung der Teilnahmeeignung.</li> <li>○ Jeder größere chirurgische Eingriff &lt; 14 Tage vor Beginn des ISM-SAF zur Testung der Teilnahmeeignung.</li> </ul> </li> <li>• Begleittherapie mit einem starken Cytochrom P450 3A4 (CYP3A4) Inhibitor oder starken oder moderaten Induktor.</li> <li>• Vorgeschichte einer bösartigen Erkrankung, die innerhalb von 3 Jahren vor der ersten Dosis des Studienmedikaments diagnostiziert wurde oder eine Therapie erforderte.</li> <li>• QT-Intervall (QTcF) &gt; 480 msec.</li> <li>• Anfallsleiden in der Vorgeschichte (z. B. Epilepsie).</li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>• Vorangegangener zerebrovaskulärer Unfall oder eine transitorische ischämische Attacke innerhalb von 12 Monaten vor der ersten Dosis des Studienmedikaments.</li> <li>• Bekanntes Risiko oder ein kürzlich vorangegangenes Ereignis (12 Monate vor erster Dosis des Studienmedikaments) einer intrakraniellen Blutung (z. B. Hirnaneurysma).</li> <li>• Klinisch signifikante, unkontrollierte kardiovaskuläre Erkrankung, einschl. einer Herzinsuffizienz Grad III oder IV nach der Klassifikation II der New York Heart Association (NYHA), einen Myokardinfarkt oder eine instabile Angina pectoris innerhalb der letzten 6 Monate, klinisch signifikante, unkontrollierte Herzrhythmusstörungen oder einen unkontrollierten Bluthochdruck.</li> </ul>
<b>Intervention und Zahl der Patientinnen und Patienten</b>	<p><b>Teil 1: RCT-Phase, mind. 12 Wochen (nicht relevant für Nutzenbewertung):</b>  Gescreent gesamt: N = k. A.  Randomisiert/Eingeschlossen: N = 39  25 mg Avapritinib: N =10  50 mg Avapritinib: N =10  100 mg Avapritinib: N =10  Placebo: N = 9</p> <p><b>Teil 2: RCT-Phase, mind. 24 Wochen (relevant für Nutzenbewertung):</b>  Gescreent gesamt: N = k. A.  Randomisiert/Eingeschlossen: N = 212  25 mg Avapritinib: N = 141  Placebo: N = 71</p> <p><b>Teil 3: Einarmige Langzeitbeobachtung (nicht relevant für Nutzenbewertung):</b>  25 mg Avapritinib: N = 235 (aus Teil 1 und 2)</p>
<b>Ort und Zeitraum der Durchführung; Datenschnitte</b>	<p><b>Studienzentren</b>  42 Studienzentren: 19 in Nordamerika, 20 in Europa und 3 in Großbritannien.</p> <p><b>Studienzeitraum</b></p> <p><u>Teil 1</u> (RCT-Phase, mind. 12. Wochen, nicht relevant für Nutzenbewertung)</p> <ul style="list-style-type: none"> <li>• Erste Person gescreent: 11.01.2019</li> <li>• Erste Person, erste Visite: 16.04.2019</li> <li>• Datenschnitt für Teil 1: 27.12.2019, geplante Interimsanalyse (12 Wochen)</li> <li>• Data-Cut-off: 31.3.2020</li> </ul> <p><u>Teil 2</u> (RCT-Phase, mind. 24 Wochen, relevant für Nutzenbewertung)</p> <ul style="list-style-type: none"> <li>• Erste Person gescreent: 23.07.2020</li> <li>• Letzte Person, erste Visite: 06.01.2022</li> <li>• Erste Person eingeschrieben: 24.09.2020</li> <li>• Datenschnitt für Teil 2 und 3: 23.06.2022 (präspezifizierte Analyse, zulassungsrelevant)</li> <li>• Datenschnitt für Teil 3: 07.04.2023 (zulassungsrelevante Analyse)</li> </ul>
<b>Endpunkte der RCT-Phase (Teil 2) gemäß Studienprotokoll</b>	<p><b>Primärer Endpunkt</b>  Mittlere Veränderung des ISM-SAF TSS von Baseline bis Woche 24<sup>3</sup>.</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Anteil der Personen mit einer <math>\geq 50\%</math>igen Reduktion der Serumtrypsin-Konzentration</li> <li>• Anteil der Personen mit einer <math>\geq 50\%</math>igen Reduktion des Anteils der KIT-D816V-Allele im peripheren Blut, oder nicht nachweisbar (<math>&lt; 0,02\%</math>) bei Personen mit nachweisbarer Mutation zu Baseline.</li> <li>• Anteil der Personen mit einer Reduktion des ISM-SAF TSS um <math>\geq 30\%</math> und <math>\geq 50\%</math>.</li> <li>• Anteil der Personen mit einer <math>\geq 50\%</math>igen Reduktion der Mastzellen im Knochenmark, oder ohne Aggregate bei Personen mit Aggregaten zu Baseline.</li> </ul>

Charakteristikum	Beschreibung
	<p><b>Weitere sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Veränderung der Krankheitslast der Mastzellen in jeder Behandlungskohorte anhand folgender Messgrößen: <ul style="list-style-type: none"> <li>○ Serumtryptase-Konzentration</li> <li>○ KIT-D816V-Allelbelastung im Blut</li> <li>○ Mastzellen im Knochenmark</li> </ul> </li> <li>• Veränderung im Gebrauch von BSC.</li> <li>• Veränderungen in den Domänen und einzelnen Symptomen des ISM-SAF (inkl. (schwerstes) Leitsymptom und (schwerste/s) „Leitdomäne/-symptomcluster“).</li> <li>• Veränderungen bei den patientenberichteten Endpunkten „MC-QoL“, „PGIS“, „SF-12“, „PGIC“ und „EQ-5D-5L“.</li> <li>• Bewertung der Sicherheit und Verträglichkeit von Avapritinib anhand von UE, Vital-, EKG- und Laborwerten.</li> <li>• Bestimmung der Pharmakokinetik von Avapritinib.</li> </ul> <p><b>Explorative Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Veränderung der Knochendichte (optional, wird nach Ermessen des Prüfpersonals durchgeführt).</li> <li>• Veränderung des prozentualen Anteils der von Mastozytose betroffenen Körperoberfläche in der Haut, Veränderung der Anzahl der Läsionen und Veränderung der Pigmentierung der Hautläsionen bei Personen mit Mastozytose in der Haut zu Baseline.</li> <li>• Veränderung der Mastzellen in der Haut bei Personen mit Mastozytose in der Haut zu Baseline.</li> <li>• Veränderung der Anzahl der Anaphylaxie-Episoden, basierend auf der Verwendung von Epinephrin.</li> </ul>

<sup>1)</sup> Da der ISM-SAF TSS während des Screenings zwei Mal über eine 14-tägige Periode erhoben wird (Eignung und Baseline), kann der Baseline-Wert unter dem Eignungswert liegen (< 28).

<sup>2)</sup> In Teil 1 lag die mediane Behandlungsdauer bei 18,15 bzw. 19,7 Wochen, allerdings wurden die relevanten Effektivitätspunkte nur bis Zyklus 4 Tag 1 (12 Wochen) ausgewertet.

<sup>3)</sup> Gemäß SAP-Analyse der ITT-Population, Auswertung der PP-Population als Sensitivitätsanalyse. In Modul 4 wurden allerdings nur Analysen zur PP-Population dargestellt.

Abkürzungen: ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; BSC: Best Supportive Care; ECOG-PS: Eastern Cooperative Oncology Group Performance Status; EQ-5D: European Quality of Life 5-Dimension 5-Level; ISM: Indolente systemische Mastozytose; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; ITT: Intention-to-Treat; k. A.: keine Angabe; LLN: Untere Normgrenze; MC-QoL: Mastocytosis Quality of Life Questionnaire; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PP: Per Protokoll; PUVA: Psoralenen mit UV-A (langwelliges UV-Licht); RCT: Randomisierte kontrollierte Studie; SAP: Statistischer Analyseplan; SF-12: Short Form 12 Health Survey; SM: Systemische Mastozytose; TSS: Gesamt-Symptomscore; UE: Unerwünschtes Ereignis; ULN: Obere Normgrenze.

Die Teilnehmenden sollten aus einem der folgenden Gründe von der Studie ausgeschlossen werden: Intrakranielle Blutungen (beliebigen Grades), anhaltende Neutropenie, Anämie oder Thrombozytopenie Grad 3 oder 4, unerwünschte Ereignisse, klinisches Fortschreiten der systemischen Mastozytose, Beginn einer neuen Therapie, Tod, Verlust der Nachbeobachtung, Protokollabweichung, Rücknahme der Einwilligungserklärung, Schwangerschaft oder Stillzeit, Entscheidung des Prüfpersonals, Non-Compliance, Entscheidung des Sponsors. Klinisches Fortschreiten (Progression) ist definiert als Entwicklung eines assoziierten hämatologischen Neoplasmas, Fortschreiten zur Mastzellleukämie, Fortschreiten zur aggressiven systemischen Mastozytose mit neuen C-Befunden gemäß WHO, die auf systemische Mastozytose zurückzuführen sind.

### Protokolländerungen

Es wurden 10 Änderungen des Originalprotokolls vom 15.02.2018 vorgenommen, davon waren 5 länderspezifische Amendments und 5 globale Amendments. Nach Einschreibung der ersten Person in Teil 2 der Studie PIONEER am 24.09.2020 wurden 2 Amendments vorgenommen.

*Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie PIONEER*

<b>Amendment</b>	<b>Wesentliche Änderungen</b>
Version 4 vom 15.06.2020	Ausschluss von Patientinnen und Patienten mit schwelender systemischen Mastozytose.
Version 7 vom 01.11.2021	Klärung der Verwendung von Kortikosteroiden während der Studie: Wenn eine Person zu Beginn der Studie keine Kortikosteroide erhielt, wurden Kortikosteroide während der Studie nicht empfohlen. Ebenso wurde für Personen, die zu Studienbeginn Kortikosteroide erhielten, eine Erhöhung der Dosierung dieser Medikamente während der Studie nicht empfohlen.
Version 9 vom 07.06.2022	Basierend auf der im Mai 2022 eingegangenen Rückmeldungen der FDA wurden die folgenden Änderungen vorgenommen: <ul style="list-style-type: none"> <li>• Änderung des primären Endpunkts für Teil 2 der Studie von "Anteil der Patienten mit einer <math>\geq 30\%</math>igen Reduktion im TSS" in "Mittlere Veränderung des ISM-SAF TSS".</li> <li>• Ergänzen des Endpunkts "Anteil der Patienten mit einer <math>\geq 30\%</math>igen Reduktion des TSS" zu den zentralen, sekundären Endpunkten für Teil 2 sowie Hinzufügen des Endpunkts "Anteil der Patienten mit einer <math>\geq 50\%</math>igen Reduktion des TSS".</li> <li>• Änderung der Analyseverfahren für die "Mittlere Veränderung des ISM-SAF TSS" von einem Zwei-Stichproben-t-Test auf eine ANCOVA, kontrolliert für den Randomisierungs-Stratifizierungsfaktor (Serumtryptase) und den Baseline-ISM-Status (mittelschwer vs. schwer).</li> <li>• Entfernung der LOCF-Methode zur Imputation fehlender Daten.</li> </ul>

Abkürzungen: ANCOVA: Kovarianzanalyse; FDA: U.S. Food and Drug Administration; ISM: Indolente systemische Mastozytose; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; LOCF: Last Observation Carried Forward; TSS: Gesamt-Symptomscore.

## Charakterisierung der Intervention der 24-wöchigen RCT-Phase (Teil 2)

Tabelle 4: Charakterisierung der Intervention Avapritinib

Intervention	Kontrolle
<p><b>Avapritinib</b> Dosierung: 25 mg Avapritinib, oral (auf nüchternen Magen), täglich in kontinuierlichen 28-Tage-Zyklen über mind. 6 Zyklen (24 Wochen). <b>+ BSC (s. u.)</b></p>	<p><b>Placebo</b> Oral, täglich in kontinuierlichen 28-Tage-Zyklen. <b>+ BSC (s. u.)</b></p>
<ul style="list-style-type: none"> <li>• Eine vorübergehende Unterbrechung des Studienmedikaments (bis zu 8 Wochen) ist für Personen zulässig, die während des Behandlungszeitraums eine Unterbrechung benötigen (z. B. für eine Operation oder ein anderes Verfahren (mit Ausnahme von Knochenmark- und Hautbiopsien) oder wegen Toxizität). Längere Unterbrechungen können mit schriftlicher Genehmigung des Sponsors erfolgen.</li> <li>• Dosisescalationen sind nicht erlaubt.</li> <li>• Dosisreduzierung (Einnahme jeden 2. Tag) und -unterbrechung ist bei UE und Thrombozytopenie erlaubt.</li> <li>• Dosisreescalationen sind erlaubt.</li> <li>• Behandlungsende bei inakzeptabler Toxizität, Fortschreiten der Erkrankung, Tod, Beendigung der Studie oder Rücktritt der Patientin / des Patienten, nach maximal 5 Jahren (bei Übergang zur einarmigen Open-label-Phase).</li> </ul>	
<p><b>Erlaubte Begleitmedikation inkl. BSC</b> Die Patientinnen und Patienten erhielten während der gesamten Studie eine Behandlung mit BSC. Dazu gehörten H1- und H2-Blocker, Protonenpumpenhemmer, Osteoklastenhemmer, Leukotrienhemmer, Kortikosteroide (in einer Dosis von höchstens 20 mg Prednison oder einem Äquivalent), Cromoglicinsäure und andere Mastzellstabilisatoren sowie dem Antiimmunglobulin-E-Antikörper Omalizumab. Dosierungen und Zeitpläne der BSC zur Behandlung der SM-Symptome mussten vor Initiierung der ISM-SAF-TSS-Eignung mindestens 14 Tage lang stabil sein (dies konnte schon vor Unterzeichnung der Einwilligungserklärung erreicht worden sein). Anzahl, Häufigkeit und Dosierung der BSC-Therapie konnte jedoch im Verlauf der Studie nach Ermessen des behandelnden Arztes geändert werden. Personen mit Osteoporose sollten zur Verbesserung der Knochendichte mit einem Bisphosphonat oder einer anderen Therapie behandelt werden.</p> <p><b>Nicht erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Jegliche antineoplastische Behandlung (einschl. Kinaseinhibitoren und Chemotherapie) mit Ausnahme des Studienmedikaments und Strahlentherapie.</li> <li>• PUVA-Therapie.</li> <li>• Jeder hämatopoetische Wachstumsfaktor.</li> <li>• Systemische Kortikosteroide (&gt; 20 mg Prednison oder Äquivalent), topische Kortikosteroide, orale Kortikosteroide und lokale Kortikosteroid-Injektionen sollten während der Studie nicht begonnen werden. Wenn eine Person bei Studienbeginn bereits Kortikosteroide erhält, wird empfohlen, die Dosierung während der Studie nicht zu erhöhen.</li> </ul>	

Abkürzungen: BSC: Best Supportive Care; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; PUVA: Psoralenen mit UV-A (langwelliges UV-Licht); SM: Systemische Mastozytose; TSS: Gesamt-Symptomscore; UE: Unerwünschtes Ereignis.

## 2.3 Endpunkte

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

Tabelle 5: Zusammenfassung der Endpunktbewertung der Studie PIONEER

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Gesamtüberleben	Mortalität	Ja <sup>5)</sup>	Ja
ISM-SAF <sup>1)</sup>	Morbidität	Ja	Ja <sup>2)</sup>
PGIS		Ja	Ja <sup>2)</sup>
Serumtryptase-Konzentration		Ja	Nein
KIT-D816V-Mutationslast		Ja	Nein
Mastzellen im Knochenmark		Ja	Nein
EQ-5D-VAS		Ja	Ja <sup>2)</sup>
PGIC		Ja	Ergänzend <sup>3)</sup>
Allgemeiner BSC-Gebrauch		Ja	Nein
Anaphylaxie-Episoden mit Epinephrin-Einsatz		Nein	Ja
Mastozytose auf der Haut		Ja	Nein
SF-12	Lebensqualität	Ja	Ja <sup>4)</sup>
MC-QoL		Ja	Ja
Unerwünschte Ereignisse	Sicherheit	Ja	Ja <sup>4)</sup>

<sup>1)</sup> Primärer Endpunkt.

<sup>2)</sup> Darstellung der mittleren Veränderung der ITT-Population in der Nutzenbewertung, ergänzende Darstellung der Responderanalyse zur PP-Population im Anhang, Begründung siehe Kapitel 2.4.

<sup>3)</sup> Ergänzende Darstellung des Endpunkts im Anhang: Mittlere Veränderung der ITT-Population und Responderanalyse der PP-Population.

<sup>4)</sup> Darstellung der mittleren Veränderung der ITT-Population in der Nutzenbewertung, ergänzende Darstellung der mittleren Veränderung (mit Effektschätzer) der PP-Population im Anhang, Begründung siehe Kapitel 2.4.

<sup>5)</sup> Wurde im Rahmen der Sicherheit erhoben.

Abkürzungen: BSC: Best Supportive Care; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; ITT: Intention-to-Treat; MC-QoL: Mastocytosis Quality of Life Questionnaire; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PP: Per Protokoll; pU: pharmazeutischer Unternehmer; SF-12: Short-Form 12 Health Survey.

### 2.3.1 Mortalität

#### **Gesamtüberleben**

Der Endpunkt „Gesamtüberleben“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Todesfälle werden im Rahmen der Beobachtung der Sicherheitsendpunkte von der Studieneinwilligung bis zum Sicherheits-Follow-up erhoben.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz und Validität

Todesfälle stellen einen patientenrelevanten Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) dar. Die Erhebung des Endpunkts „Gesamtüberleben“ wird als valide erachtet.

### 2.3.2 Morbidität

#### **Symptome mittels „ISM-SAF“ und „PGIS“**

##### **ISM-SAF**

Der Endpunkt „ISM-SAF“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der „Indolent Systemic Mastocytosis Symptom Assessment Form“ (ISM-SAF) ist ein patientenberichteter Fragebogen zur Erfassung der Symptome von Personen mit ISM. Er umfasst 11 Symptom-Items, die mittels einer 11-Punkt-Skala (0–10) in Bezug auf den Schweregrad bewertet werden sollen, sowie ein Item zur Durchfallhäufigkeit. Das Instrument soll täglich in Form eines elektronischen Tagebuchs ausgefüllt werden. Die Items werden verschiedenen Domänen zugeordnet (siehe Tabelle 6). Die Domänenscores können Werte zwischen 0 und 30 erreichen, der Gesamt-Symptomscore (TSS) Werte zwischen 0 und 110. Je höher die Punktezahl ist, desto stärker sind die Symptome ausgeprägt. Bei dem Item „Durchfallhäufigkeit“ sollen die Patientinnen und Patienten einen Zahlenwert eingeben.

In der Studie PIONEER sollten neben dem Gesamtscore, den Domänenscores und den Items die Werte des (schwersten) Leitsymptoms und der/des (schwersten) Leitdomäne/-symptomclusters verglichen werden. Das Leitsymptom ist das Symptom mit dem höchsten Score zu Baseline bei jeder Person. Der Wert zu Zyklus 7 Tag 1 des Leitsymptoms bezieht sich auf den Wert für dieses Symptom. Für Personen mit mehreren Leitsymptomen oder Leitdomänen/-symptomclustern, wird der Durchschnitt der Veränderung zu Baseline berechnet.



Tabelle 6: Items und Domänen des ISM-SAF

Item	Domänenscore	Gesamtscore
Bauchschmerzen	Gastrointestinaler Symptomscore (GSS)	Gesamt-Symptomscore (TSS)
Übelkeit		
Durchfall / Diarrhö		
Flecken auf der Haut	Haut-Symptomscore (SSS)	
Juckreiz		
Hitzewallungen		
Knochenschmerzen	Knochenschmerzen	
Erschöpfung	Erschöpfung	
Schwindelgefühl	Neurokognitives Symptomcluster (NSS)	
Gehirnnebel		
Kopfschmerzen		
Durchfall	Durchfallhäufigkeit	

Abkürzungen: ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form.

Zur Bestimmung des Baseline-Werts füllten die Patientinnen und Patienten den ISM-SAF in den 14 Tagen vor der ersten Studiendosis täglich aus. Anschließend wurde der Mittelwert gebildet. Auch während der Behandlungsphase wird der ISM-SAF täglich erhoben und jeweils der 14-Tage-Durchschnitt vor der Visite berechnet.

Für die Berechnung des 14-Tage-Durchschnitts mussten mindestens 7 Tages-Scores vorhanden sein. Personen, bei denen entweder der Baseline-Score oder der Wert zu Zyklus 7 Tag 1 fehlen, wurden von der Analyse ausgeschlossen. Wenn die fehlenden Daten völlig zufällig fehlen („Missing Completely at Random“), hat die gesamte Fallanalyse aufgrund des geringeren Stichprobenumfangs gemäß SAP eine geringere statistische Aussagekraft, aber die beobachteten Daten sind nicht verzerrt.

### *Bewertung*

Die Operationalisierung ist weitgehend nachvollziehbar, das Instrument wurde vorgelegt.

### Patientenrelevanz

Der Endpunkt „ISM-SAF“ wird als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

### Validität

Es liegen eine Studie zur Fragebogenentwicklung sowie zwei Validierungsstudien vor, in die Patientinnen und Patienten mit ISM einschlossen wurden [12,13,15]. Das im Studienbericht zitierte ISM-SAF Evidence Dossier wurde nicht vorgelegt.

In die Entwicklung des Fragebogens wurden Patientinnen und Patienten mit ISM einbezogen, dies kann als ausreichender Hinweis auf Inhaltsvalidität gesehen werden [15]. Die interne Konsistenz ist für den Gesamtscore sowie die Domänenscores „Gastrointestinaler Symptomscore“ (GSS) und „Haut-Symptomscore“ (SSS) gut (Cronbachs Alpha > 0,7). Die Items korrelierten mäßig bis stark mit anderen Items der jeweiligen Domäne und wenig bis mäßig mit den Items anderer Domänen. Analysen zur konvergenten und Known-Group-Validität zeigten, dass die Scores des ISM-SAF stark mit den entsprechenden Scores des SF-12, MC-QoL und PGIS korrelierten, wobei der TSS stärker mit den anderen patientenberichteten Endpunkten (PRO) korreliert als die Domänenscores GSS und SSS. Beim Vergleich der Mittelwerte von zwei aufeinanderfolgenden Wochen zeigte sich eine

hohe Test-Retest-Reliabilität für den Gesamt- und die Domänenscores ( $ICC > 0,93$ ;  $n = 61$ ). Es konnte eine Veränderung der Symptome mit einer moderaten Effektstärke aufgezeigt werden ( $0,8 > d > 0,5$ ). Dies wird als Nachweis für die Änderungssensitivität angesehen. [12,13]

Es ist anzumerken, dass keine Informationen zu den psychometrischen Eigenschaften des neurokognitiven Symptomclusters (NSS) vorliegen.

TSS, GSS und SSS werden als ausreichend valide angesehen. Das NSS hingegen ist nicht ausreichend validiert und wird daher ergänzend dargestellt.

### **PGIS**

Der Endpunkt „PGIS“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der „Patient Global Impression of Severity“ (PGIS) ist eine patientenberichtete Skala zur Erfassung der Schwere einer Erkrankung. Sie besteht aus folgendem Item „Die Symptome meiner systemischen Mastozytose sind im Moment...“ und die Einschätzung erfolgt auf einer 5-stufigen Skala:

- 0 = nicht vorhanden: Keine Symptome.
- 1 = minimal: Symptome lassen sich leicht ignorieren.
- 2 = moderat: Symptome lassen sich schwer ignorieren.
- 3 = stark: Symptome lassen sich sehr schwer ignorieren.
- 4 = sehr stark: Symptome lassen sich nicht ignorieren.

Der PGIS wird zusammen mit dem MC-QoL, PGIC, SF-12 und der EQ-5D-VAS bei jedem Studienbesuch bis zum Behandlungsende ausgefüllt. Die PRO werden mit einem elektronischen Tablet erhoben, in Ausnahmefällen ist die Erhebung auch in Papierform möglich.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar, jedoch fehlen Informationen zur konkreten Administration des Instruments in der Studie (u. a. Zeitpunkt der Erhebung innerhalb einer Studienvsiste). Das in der Studie eingesetzte Instrument wurde vorgelegt.

#### Patientenrelevanz und Validität

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Die Einschätzung der Patientinnen und Patienten zum Schweregrad der jeweils eigenen erkrankungsspezifischen Symptomatik ist als patientenrelevant zu bewerten, wobei eine Erhebung per PGIS grundsätzlich als geeignet eingeschätzt wird. Mit dem Instrument wird die Symptomschwere zum Zeitpunkt der Visite bewertet. Es ist einschränkend zu berücksichtigen, dass keine Überprüfung der Gütekriterien des Items und der Skalenausprägungen (0–4) bezüglich der systemischen Mastozytose vorliegt.

### **Mastzellen im Knochenmark, Serumtrypsin-Konzentration, KIT-D816V-Mutationslast**

Die Endpunkte „Mastzellen im Knochenmark“, „Serumtrypsin-Konzentration“ und „KIT-D816V-Mutationslast“ werden in der Nutzenbewertung aufgrund nicht-gegebener Patientenrelevanz nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Die laborparametrische Symptomatik wird in der Studie PIONEER anhand folgender Verfahren erhoben:

- Knochenbiopsie und -aspirat (Endpunkt: „Mastzellen im Knochenmark“)
- Blutproben (Endpunkte: „Serumtrypsin-Konzentration“, KIT-D816V-Mutationslast)

#### Patientenrelevanz

Bei den Endpunkten handelt es sich um Laborparameter, sodass die Erhebung der Morbidität nicht primär anhand von Krankheitssymptomen, sondern auf Basis von asymptomatischen, nicht unmittelbar patientenrelevanten Befunden geschieht. Es wurde keine Validierung als Surrogat vorgelegt.

#### Validität

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

### **Allgemeiner Gesundheitszustand mittels „EQ-5D-VAS“ und „PGIC“**

#### **EQ-5D-VAS**

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

#### Operationalisierung

##### *Beschreibung*

Zur Erfassung des allgemeinen Gesundheitszustands wurde der „European Quality of Life 5-Dimension 5-Level“ (EQ-5D) erhoben; in Modul 4 wurde die visuelle Analogskala des EQ-5D (EQ-5D-VAS) dargestellt. Die EQ-5D-VAS ist eine Skala von 0 bis 100, auf der die Patientinnen und Patienten ihren Gesundheitszustand einschätzen. Ein Wert von 0 entspricht dem denkbar schlechtesten Gesundheitszustand, ein Wert von 100 dem denkbar besten Gesundheitszustand. Die EQ-5D-VAS wird zusammen mit dem MC-QoL, PGIS, PGIC und SF-12 bei jedem Studienbesuch bis zum Behandlungsende ausgefüllt. Die PRO werden mit einem elektronischen Tablet erhoben, in Ausnahmefällen ist die Erhebung auch in Papierform möglich.

##### *Bewertung*

Die Operationalisierung ist eingeschränkt nachvollziehbar, denn es fehlen Informationen zur konkreten Administration des Instruments in der Studie (u. a. Zeitpunkt der Erhebung innerhalb einer Studienvisite). Das in der Studie eingesetzte Instrument wurde vorgelegt.

#### Patientenrelevanz

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Die EQ-5D-VAS bildet den allgemeinen Gesundheitszustand ab, wird für die Nutzenbewertung herangezogen und der Endpunktkategorie „Morbidität“ zugeordnet.

### Validität

Die Validität der EQ-5D-VAS wurde in vorherigen Nutzenbewertungsverfahren bereits in verschiedenen Anwendungsgebieten bewertet und das Instrument wird auch im zugrundeliegenden Anwendungsgebiet als valide eingeschätzt.

### **PGIC**

Beim Endpunkt „PGIC“ wird die zweite Frage zum Grad der Veränderung aufgrund eingeschränkter Validität im Anhang dargestellt.

### Operationalisierung

#### *Beschreibung*

In der Studie PIONEER sollten die Patientinnen und Patienten im Rahmen der Erhebung des „Patient Global Impression of Change“ (PGIC) zwei Single-Item-Skalen ausfüllen. Bei dem ersten Item handelt es sich um eine 7-Punkte-Skala, die die Einschätzung der Patientin / des Patienten zur allgemeinen Verbesserung darstellt, mit dem zweiten Item wird der Grad der Veränderung seit Beginn der Behandlung auf einer Skala von 0 bis 10 bewertet.

Frage 1 lautet: „Seit der Behandlung in dieser Klinik, wie würdest du die Veränderung (wenn es eine gab) in Aktivitätseinschränkungen, Symptomen, Emotionen und Lebensqualität beschreiben, bezogen auf deinen schmerzhaften Zustand?“

Folgende Antwortmöglichkeiten standen zur Auswahl: (1) Keine Verbesserung (oder der Zustand wurde schlechter); (2) Fast der gleiche Zustand; (3) Ein bisschen besser, aber keine bemerkenswerte Verbesserung; (4) Etwas besser, aber die Änderung hat keinen wirklichen Unterschied gemacht; (5) Mäßig besser, und eine leichte, aber spürbare Veränderung; (6) Besser und eine eindeutige Verbesserung, die einen echten und lohnenden Unterschied gemacht hat; (7) Sehr viel besser und eine erhebliche Verbesserung, die den großen Unterschied ausgemacht hat.

Frage 2 lautet: „Bitte kreuzen Sie unten [auf der visuellen Analogskala] die Zahl an, die dem Grad Ihrer Veränderung seit Beginn der Behandlung in dieser Klinik entspricht“. Die Skala geht von 0 (viel besser) über 5 (keine Veränderung) bis 10 (viel schlechter).

Auf der 7-Punkte-Skala zeigen höhere Werte eine stärkere Verbesserung der Symptome an, auf der 11-Punkte-Skala niedrigere Werte.

Der PGIC wird zusammen mit dem MC-QoL, PGIS, SF-12 und der EQ-5D-VAS bei jedem Studienbesuch bis zum Behandlungsende ausgefüllt. Zu Baseline fand keine Erhebung statt. Die PRO werden mit einem elektronischen Tablet erhoben, in Ausnahmefällen ist die Erhebung auch in Papierform möglich.

#### *Bewertung*

Die Operationalisierung ist eingeschränkt nachvollziehbar, denn es fehlen Informationen zur konkreten Administration des Instruments in der Studie (u. a. Zeitpunkt der Erhebung innerhalb einer Studienvisite). Das in der Studie eingesetzte Instrument wurde vorgelegt.

### Patientenrelevanz

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

### Validität

Die eingesetzte Version des PGIC entspricht nicht der herkömmlichen Version des Instruments. Bei der ersten Frage ist unklar, ob Patienten in der Lage sind, die Verbesserung anhand einer 7-Punkte-Skala zu bewerten, da sich die Antwortmöglichkeiten sehr ähneln. Darüber hinaus gibt es nur eine Antwortmöglichkeit für „Keine Veränderung oder Verschlechterung“, sodass nicht erkennbar ist, bei wie viel Patienten eine Verschlechterung eingetreten ist. Bei der visuellen Analogskala der zweiten Frage befindet sich die Antwortmöglichkeit „keine Veränderung“ in der Mitte der Skala, „Verbesserung“ und „Verschlechterung“ an den beiden Extrema. Auf diese Weise ist für den Patienten und Auswerter leicht erkennbar, ob eine Verbesserung oder Verschlechterung eingetreten ist und wie groß diese war. Aus diesem Grund wird die zweite Frage gegenüber der ersten bevorzugt und zur Nutzenbewertung herangezogen.

Bei der rückwirkenden Einschätzung einer Veränderung müssen die Befragten in der Lage sein, sowohl ihren gegenwärtigen Zustand als auch ihren Ausgangszustand zu quantifizieren und daraus eine mentale Subtraktion durchzuführen. Hier ist kritisch anzumerken, dass ein möglicher „Recall Bias“ nicht auszuschließen ist. Stattdessen wird bei dieser Abschätzung größtenteils der gegenwärtige Gesundheitszustand erfasst. Dieser Effekt wird umso größer, je länger das Erinnerungsintervall ist [8,9,10,11]. Daher werden die Ergebnisse der zweiten Frage im Anhang dargestellt.

### **Allgemeiner BSC-Gebrauch**

Der Endpunkt „Allgemeiner BSC-Gebrauch“ wird aufgrund nicht-gegebener Patientenrelevanz nicht berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Der Endpunkt „Allgemeiner BSC-Gebrauch“ ist definiert als der Anteil an Personen, die zwischen Baseline und Woche 24 eine Veränderung (Verringerung) der Best-Supportive-Care (BSC)-Medikation hatte.

Zu den allgemeinen BSC-Medikamenten gehören H1- und H2-Blocker, Protonenpumpenhemmer, Leukotrienhemmer, Kortikosteroide (die Dosis darf 20 mg Prednison oder ein Äquivalent nicht überschreiten, andernfalls können die Teilnehmenden als Non-Responder eingestuft werden), Cromoglicinsäure, Omalizumab und Osteoklastenhemmer.

Im Rahmen des Screenings wurde bei den Patientinnen und Patienten die BSC-Medikation zur Behandlung der Symptome der systemischen Mastozytose optimiert und stabilisiert. In der Behandlungsphase mit dem Studienmedikament konnte auf Grundlage der vom Prüfpersonal beurteilten Symptome die Dosis schrittweise reduziert werden. Dosierung und Zeitplan aller BSC-Medikamente wurden während des Screenings und der Behandlungsphasen in der elektronischen Patientenakte erfasst. Die letzte Erhebung erfolgte 30 Tage nach der letzten Dosis im Rahmen der Sicherheitsbeobachtung.

#### *Bewertung*

Die Operationalisierung ist überwiegend nachvollziehbar.

### Patientenrelevanz

Der Endpunkt „Allgemeiner BSC-Gebrauch“ bildet Dosen und/oder Applikationshäufigkeiten ab, welche nicht per se patientenrelevant sind. Sollten diesbezügliche Elemente relevant für Patientinnen / Patienten sein, so sollte sich dies in den Endpunktkategorien „Sicherheit“, „Lebensqualität“ und/oder „Morbidity“ widerspiegeln.

### Validität

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

### **Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin**

Der Endpunkt „Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Beim Endpunkt „Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin“ wird die Anzahl an Personen dargestellt, bei denen eine Epinephrin-Gabe aufgrund von anaphylaktischen Episoden erforderlich war.

Die Begleitmedikation einschließlich Epinephrin sollte ab Unterzeichnung der Einverständniserklärung in der elektronischen Patientenakte aufgezeichnet werden. Die letzte Erhebung erfolgte 30 Tage nach der letzten Dosis im Rahmen der Sicherheitsbeobachtung.

##### *Bewertung*

Die Operationalisierung ist überwiegend nachvollziehbar. Die Analyse war präspezifiziert.

#### Patientenrelevanz

Entsprechend § 2 Satz 3 AM-NutzenV ist der Nutzen eines Arzneimittels „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Anaphylaxie-Episoden, die den Einsatz von Epinephrin erfordern, bilden im vorliegenden Anwendungsgebiet die Morbidität ab und sind patientenrelevant.

#### Validität

Da im Anwendungsgebiet Epinephrin insbesondere bei anaphylaktischen Episoden eingesetzt wird, eignet sich die Messung von Epinephrin-Gaben, um starke anaphylaktische Episoden abzubilden.

### **Mastozytose auf der Haut**

Der Endpunkt „Mastozytose auf der Haut“ wird aufgrund unklarer Patientenrelevanz und unvollständiger Erhebung innerhalb der Studienpopulation in der Nutzenbewertung nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Folgende Operationalisierungen wurden für den Endpunkt „Mastozytose auf der Haut“ im SAP beschrieben:

- Veränderung der Pigmentierung (Hautfarbe) der Läsionen, dargestellt nach betroffenem Körperteil (am stärksten betroffenes Körperteil, hinterer Oberschenkel, hinterer Torso, vorderer Oberschenkel, vorderer Torso) als Wasserfalldiagramm.
- Veränderung der von der Mastozytose betroffenen Fläche, dargestellt nach betroffenem Körperteil als Wasserfalldiagramm.
- Maximale Reduktion der betroffenen Fläche, dargestellt als Wasserfalldiagramm.

Bei Personen, die beim Screening eine Mastozytose in der Haut aufwiesen, konnte in der randomisiert-kontrollierten Studienphase (RCT-Phase, Teil 2) während des Screenings sowie nach 3 und 6 Monaten eine standardisierte Fotografie der Haut gemacht werden. Diese wurde vor der Hautbiopsie durchgeführt. Sie war freiwillig und der Patient konnte sich jederzeit dagegen entscheiden. Details zur standardisierten Fotografie werden laut Protokoll in der Studienanleitung (Study Manual) beschrieben, dieses liegt allerdings nicht vor.

Im Studienbericht wird beschrieben, dass die Fotos mit der gleichen Kameraausrüstung aufgenommen wurden. Das Personal wurde in der ordnungsgemäßen Verwendung der Ausrüstung und

der Erstellung standardisierter Fotos geschult. Bei jedem Bild wurde von einem Algorithmus zur Läsionserkennung die Anzahl der Läsionen und der Anteil der Fläche bestimmt, die von kutaner Mastozytose betroffen war. Die vom Algorithmus bewerteten Bilder wurden außerdem von einem Bildanalysetechniker visuell bewertet. Abschließend wurden die Fotografien von einem verblindeten Hautbewertungsausschuss hinsichtlich des Ansprechens auf die Behandlung beurteilt.

#### *Bewertung*

Die Operationalisierung des Endpunkts ist weitgehend nachvollziehbar. Durch die mehrfache Bewertung der Fotografien (algorithmische Bildbewertung und Beurteilung des Hautbewertungsausschusses) wurde eine objektive Bewertung der Anzahl und Größe der Läsionen sichergestellt. Es ist unklar, ob eine adäquate Differenzierung der Läsionsfarbe mit Fotografien möglich ist, und ob Kriterien zur Bewertung der Veränderung der Läsionsfarbe vorlagen. Darüber hinaus erfolgte die Erhebung des Endpunkts auf freiwilliger Basis und es nahm nur etwa die Hälfte der Patientinnen und Patienten aus beiden Studienarmen an der Erhebung teil. Somit liegt ein Randomisierungsbruch vor und es ist unklar, ob die Nichtteilnahme zufällig war und die Kriterien für „Missing Completely at Random“ erfüllt waren.

#### Patientenrelevanz

Die Patientenrelevanz des Endpunkts ist unklar. Eine Reduktion der Läsionen kann je nach Ausmaß patientenrelevant sein, wenn damit ein positiver Effekt für die Betroffenen, z. B. geringerer Juckreiz, verbunden ist. Anhand der vorliegenden Auswertung der prozentualen Veränderung lassen sich dazu jedoch keine direkten Aussagen ableiten.

#### Validität

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

### **2.3.3 Lebensqualität**

#### **Gesundheitsbezogene Lebensqualität mittels SF-12**

Der Endpunkt „SF-12“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der „Short Form 12 Health Survey“ (SF-12) dient als generischer Fragebogen zur Erfassung von Symptomen, Funktionsfähigkeit und Lebensqualität. Er ist eine Kurzform des SF-36 und beinhaltet nur 12 der 36 Items, umspannt aber trotzdem (mit jeweils einem oder zwei Einzelitems) das komplette Feld der 8 Subskalen und der beiden Domänen „Physical Health“ und „Mental Health“.

Der SF-12 besteht aus den folgenden Domänen:

- Physical Component Summary (PCS)
  - Allgemeine Gesundheitswahrnehmung (1 Item)
  - Physische Funktionsfähigkeit (2 Items)
  - Physisch-bedingte Rollenfunktion (2 Items)
  - Körperliche Schmerzen (1 Item)
  - Vitalität (1 Item)
- Mental Component Summary (MCS)
  - Mentale Gesundheit (2 Items)
  - Emotional-bedingte Rollenfunktion (2 Items)
  - Soziale Funktionsfähigkeit (1 Item)

Wie auch beim SF-36 existiert eine zweite Version des SF-12, in der im Unterschied zu Version 1 bei allen Fragen 3- bis 5-Punkte-Likert-Skalen vorgegeben sind. Die Angaben der Patientinnen und Patienten werden für die PCS und MCS auf eine Skala von 0 Punkte (schlechtester Gesundheitszustand) bis 100 Punkte (bester Gesundheitszustand) umgerechnet.

In der Studie PIONEER kommt die aktuelle Version des Fragebogens (SF-12v2) zum Einsatz. Der Fragebogen wird mit einem elektronischen Tablet erhoben, in Ausnahmefällen ist die Erhebung auch in Papierform möglich. Der SF-12 wird zusammen mit dem MC-QoL, PGIS, PGIC und der EQ-5D-VAS bei jedem Studienbesuch bis zum Behandlungsende ausgefüllt. Zur Analyse sollten die Domänenscores, der PCS und der MCS berechnet werden.

#### *Bewertung*

Die Operationalisierung ist eingeschränkt nachvollziehbar, denn es fehlen Informationen zur konkreten Administration des Instruments in der Studie (u. a. Zeitpunkt der Erhebung innerhalb einer Studienvisite). Es wurde die neueste Version des Fragebogens verwendet, das in der Studie eingesetzte Instrument wurde vorgelegt.

#### Patientenrelevanz

Die gesundheitsbezogene Lebensqualität wird in der Studie PIONEER mithilfe des SF-12 erfasst. Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

#### Validität

Der SF-12 wurde im Jahr 1996 entwickelt und steht aktuell in der überarbeiteten Version v2 zur Verfügung. Die Entwicklung auf Basis des SF-36 basiert auf einem Regressionsanalyseverfahren, um die 12 Items aus dem SF-36 auszuwählen [16].

Eine Validierung für Personen mit ISM liegt nicht vor. Die Normierung erfolgt laut SAP auf die durchschnittliche US-Population. Von einer Übertragbarkeit auf den deutschen Raum ist in begrenztem Umfang auszugehen.

Da die Validität des Endpunkts „SF-12“ bereits in vorherigen Nutzenbewertungsverfahren bewertet wurde, wird sie auch im zugrundeliegenden Anwendungsgebiet als valide eingeschätzt.

#### **Krankheitsspezifische Lebensqualität mittels MC-QoL**

Der Endpunkt „MC-QoL“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der „Mastocytosis Quality of Life Questionnaire“ (MC-QoL) ist ein krankheitsspezifischer Fragebogen zur Erfassung der Einschränkungen der gesundheitsbezogenen Lebensqualität bei erwachsenen Patientinnen und Patienten mit kutaner oder indolenter systemischer Mastozytose.

Der Fragebogen besteht aus 27 Items und umfasst folgende Domänen:

- Symptome (9 Items)
- Soziales Leben / Funktionsfähigkeit (9 Items)
- Emotionen (6 Items)
- Haut (3 Items)

Die Patientinnen und Patienten beantworten die Fragen mit Bezug auf die letzten 2 Wochen auf einer 5-Punkte-Likert-Skala.

Der Gesamtscore und die Domänenscores ergeben sich durch Addition der erreichten Item-Punktwerte und anschließender linearer Transformation auf eine Skala von 0 bis 100. Eine höhere Beeinträchtigung der Lebensqualität spiegelt sich in einem höheren Score wider.



In der Studie PIONEER wird der MC-QoL zusammen mit dem PGIS, PGIC, SF-12 und der EQ-5D-VAS bei jedem Studienbesuch bis zum Behandlungsende ausgefüllt. Die PRO werden mit einem elektronischen Tablet erhoben, in Ausnahmefällen ist die Erhebung auch in Papierform möglich.

#### *Bewertung*

Die Operationalisierung ist eingeschränkt nachvollziehbar, denn es fehlen Informationen zur konkreten Administration des Instruments in der Studie (u. a. Zeitpunkt der Erhebung innerhalb einer Studienvisite). Das in der Studie eingesetzte Instrument wurde vorgelegt.

#### Patientenrelevanz

Die gesundheitsbezogene Lebensqualität wird in der Studie PIONEER mithilfe des krankheits-spezifischen Instruments MC-QoL erfasst. Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

#### Validität

Der MC-QoL wurde unter Beteiligung von Patientinnen und Patienten mit kutaner und indolenter systemischer Mastozytose entwickelt und in dieser Population validiert (N = 158, davon 76 Personen mit ISM) [14]. Es zeigte sich eine gute interne Konsistenz und Test-Retest-Reliabilität für drei der vier Domänen und den Gesamtscore. Allerdings deuten die Ergebnisse der Validierungsstudie nur auf eine moderate interne Konsistenz und Test-Retest-Reliabilität der Domäne „Haut“ hin. Die konvergente und diskriminante Validität wurde bestätigt. Es fehlen Untersuchungen zur Änderungssensitivität. Trotz dieser Limitationen wird der Endpunkt zur Nutzenbewertung herangezogen.

### **2.3.4 Sicherheit**

#### **Unerwünschte Ereignisse**

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

#### Operationalisierung

##### *Beschreibung*

Unerwünschte Ereignisse (UE) sind in der Studie PIONEER definiert als jegliche unerwünschte medizinische Ereignisse, unabhängig davon, ob sie als arzneimittelbedingt angesehen werden oder nicht. Ein UE kann jedes ungünstige und unbeabsichtigte Anzeichen (z. B. anormaler Laborbefund), Symptom oder jede Krankheit sein, die zeitlich mit der Anwendung eines Arzneimittels zusammenhängt, ohne dass ein Urteil über die Kausalität möglich ist. Ein UE kann bei jeder Anwendung des Arzneimittels (z. B. Off-Label-Use, Anwendung in Kombination mit einem anderen Arzneimittel) und bei jedem Verabreichungsweg, jeder Zusammensetzung oder Dosis auftreten, auch bei einer Überdosierung. Als Überdosierung gilt jede Dosis, die höher ist als die vorgesehene Dosis.

Ein abnormer Laborparameter wird nur dann als UE erfasst, wenn dieser zu Behandlungsabbruch oder -unterbrechung, Dosismodifikation oder einer therapeutischen Intervention führt bzw. vom Prüfpersonal als klinisch relevant eingeschätzt wird. Eine Progression der zugrundeliegenden Erkrankung soll i. d. R. nicht als UE, schwerwiegendes unerwünschtes Ereignis (SUE) oder Todesgrund berichtet werden. Stattdessen sollen UE (oder SUE), die als Komplikationen der Krankheitsprogression bewertet werden, berichtet werden. Wenn jedoch keine spezifischen Komplikationen aufgrund einer Krankheitsprogression, welche die klinischen Beobachtungen erklären, identifiziert werden können, darf „Krankheitsprogression“ als UE, SUE oder Todesgrund berichtet werden.

In der Studie PIONEER sollten UE ab Einverständniserklärung bis 30 Tage nach der letzten Gabe der Studienmedikation erhoben werden. Die Sicherheitsnachbeobachtung, die 30 Tage nach der

letzten Studienmedikation durchgeführt wird, erfolgte telefonisch. Nach der Sicherheitsnachbeobachtung sollten weiterhin SUE erhoben werden. Es liegen keine Angaben vor, wie diese erhoben werden sollten und bis zu welchem Zeitpunkt. Im SAP vom 07.06.2022 steht ergänzend, dass für die 24-wöchige RCT-Phase (Teil 2) alle UE ab Tag 1 bis zu einem Tag vor dem Übergang in die einarmige Langzeitbeobachtung (Teil 3) erfasst werden, wenn die Person zur Langzeitbeobachtung übergeht, und bis 30 Tage nach der letzten Studienmedikation, wenn die Person nach Teil 2 die Studie beendet. UE, die zwischen der Einverständniserklärung und der ersten Einnahme der Studienmedikation auftreten, werden als Vor-Behandlung-UE gewertet.

UE wurden durch spontane Berichte, Befragungen und Untersuchungen (körperliche Untersuchung, diagnostische Verfahren) durch das Prüfpersonal erhoben und in der elektronischen Krankenakte dokumentiert.

Das Auftreten von UE wird gemäß Medical Dictionary for Regulatory Activities (MedDRA) nach Systemorganklasse und Preferred Term erfasst. Die Schweregradierung erfolgt gemäß Common Terminology Criteria for Adverse Events (CTCAE, Version 5.0) des National Cancer Institute (NCI). Jene UE, die nicht in den jeweiligen CTCAE-Versionen vorliegen, werden in den Studien anhand folgender Einteilung bewertet:

- Grad 1: mild; das Ereignis ist für die Person spürbar, beeinträchtigt aber nicht Routinetätigkeiten.
- Grad 2: moderat; das Ereignis beeinträchtigt die Routinetätigkeiten, spricht aber auf symptomatische Therapie oder Ruhe an.
- Grad 3: schwer; das Ereignis schränkt die Fähigkeit zur Ausübung von Routinetätigkeiten trotz symptomatischer Therapie spürbar ein.
- Grad 4: lebensbedrohlich; ein Ereignis, welches die Person zum Zeitpunkt des Ereignisses in Lebensgefahr bringt.
- Grad 5: tödlich; ein Ereignis, welches zum Versterben der Person führt.

SUE sind definiert als jegliches Ereignis, das eines der folgenden Kriterien aufweist:

- Tod.
- Lebensbedrohlich; d. h., wenn die Person durch das Ereignis selbst mit der vorliegenden Intensität einem direkten Todesrisiko ausgesetzt ist.
- Bedurfte einer Hospitalisierung oder der Verlängerung einer bestehenden Hospitalisierung. Ausgenommen davon sind:
  - Elektive Eingriffe und Routineuntersuchungen, die nicht Folge eines UE sind;
  - Beobachtungen und/oder Behandlungen, die auch außerhalb des Krankenhauses angemessen gewesen wären, wobei im Zweifelsfall eine Bewertung als SUE erfolgen soll;
  - persistierende oder signifikante Behinderungen/Invalidität; d. h., wenn eine substantielle und/oder permanente Störung der Fähigkeit, normale Lebensfunktionen auszuführen, gegeben ist.
- Kongenitale Anomalie / ein Geburtsfehler.
- Andere, d. h. Ereignisse, die nicht zum Tod führen, lebensbedrohlich sind oder eine Hospitalisierung erfordern, sofern gemäß adäquater medizinischer Beurteilung eine Gefährdung der Person vorliegt und ggf. einer medizinischen oder chirurgischen Intervention bedürfen, um eine dieser gelisteten Folgen zu verhindern. Dazu gehören z. B.:
  - Intrakranielle Blutungsereignisse;
  - Intensivbehandlung in einer Notaufnahme oder Zuhause aufgrund eines allergischen Bronchospasmus;
  - Blutanomalien oder Konvulsionen, welche nicht zur stationären Hospitalisierung führen;
  - Entwicklung von Arzneimittelabhängigkeit oder -missbrauch.

Alle intrakraniellen Blutungsereignisse müssen als SUE berichtet werden.

Die Erhebungszeitpunkte von UE in der Studie PIONEER sind Tabelle 7 zu entnehmen.

### Unerwünschte Ereignisse von besonderem Interesse

Im SAP der Studie PIONEER (07.06.2022) werden folgende UE von besonderem Interesse (UESI) spezifiziert:

- Kognitive Effekte (17 Preferred Terms): Gedächtnisstörung; Kognitive Störung; Verwirrtheit; Amnesie; Somnolenz; Sprachstörung; Delirium; Halluzination; Stimmungsschwankungen; Agitation; Persönlichkeitsveränderung; Demenz; Veränderungen des mentalen Status; Psychotische Störung; Desorientierung; Enzephalopathie; Psychische Beeinträchtigung.
- Intrakranielle Blutungen (4 Preferred Terms): Hirnblutung; Intrakranielle Blutung; Subdurales Hämatom; Zerebrales Hämatom.

### Ödeme – Angepasste MedDRA Queries

Im SAP vom 07.06.2022 wird beschrieben, dass Ödeme nach denselben Regeln wie UESI dargestellt werden sollen.

Zu den Ödemen zählen: Bauchdeckenödem; Kapillarlecksyndrom; Übermäßige Gewichtszunahme zwischen Dialysen; Gravitationsödem; Hydrämie; Hypervolämie; Lokalisiertes Ödem; Nicht eindrückbares Ödem; Blasenbildung durch Ödem; Schwellung um die Augenhöhle; Ödem an der Stomastelle; Schwellung des Gesichts; Schwellung des Auges; Eingeweidenödem; Ödem; Generalisiertes Ödem; Augenödem; Schwellung des Augenlids; Flüssigkeitsretention; Augenlid-ödem; Gesichtsödem; Periorbitalödem; Ödem peripher.

### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar.

### Patientenrelevanz und Validität

Das Auftreten von UE ist patientenrelevant. Die Patientenrelevanz von Laborparametern als Teil der UE ist unklar. Die Erhebung der UE wird als weitestgehend valide bewertet. Die Ergebnisse werden in der Nutzenbewertung herangezogen. Bei der Ergebnisinterpretation ist zu beachten, dass die berichteten UE zum Teil auch Ereignisse der Grunderkrankung abbilden können.

### 2.3.5 Erhebungszeitpunkte

Eine Übersicht der Erhebungszeitpunkte der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte findet sich in Tabelle 7.

*Tabelle 7: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der 24-wöchigen RCT-Phase (Teil 2) der Studie PIONEER*

Studiensite \ Endpunkt	Screening-Phase (max. 14 Wochen)	Baseline (Zyklus 1 Tag 1)	Behandlungsphase (7 Zyklen à 4 Wochen = 24 Wochen, Tag 1 von jedem Zyklus)	Behandlungsende	Safety-Follow-up
Gesamtüberleben	Erhebung im Rahmen der Sicherheit				
ISM-SAF <sup>1)</sup>	Tägliche Erhebung				
EQ-5D-VAS		X	X	X	
PGIC <sup>2)</sup>			X	X	
PGIS		X	X	X	
Anaphylaxie-Episoden mit Epinephrin-Einsatz	X				
MC-QoL		X	X	X	
SF-12		X	X	X	
Unerwünschte Ereignisse <sup>3)</sup>	X				

<sup>1)</sup> Tägliche Erhebung. Nach der Optimierung der BSC-Therapie werden die ISM-SAF-TSS-Werte der darauffolgenden 14 Tage verwendet, um die Studienteilnahme zu bestimmen. Als Baseline-Wert werden die ISM-SAF-Werte der letzten 14 Tage vor der ersten Einnahme der Studienmedikation verwendet.

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

<sup>3)</sup> SUE werden vom Screening bis zur Sicherheitsbeobachtung (30 Tage nach der letzten Dosis) erfasst. Alle SUE, die > 30 Tage nach der letzten Dosis des Studienmedikaments auftreten, werden ebenfalls gemeldet. UE, die vor der Behandlungsphase erfasst werden, werden als Vor-Behandlung-UE gewertet.

Abkürzungen: BSC: Best Supportive Care; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; MC-QoL: Mastocytosis Quality of Life Questionnaire; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; RCT: Randomisierte kontrollierte Studie; SF-12: Short Form 12 Health Survey; SUE: Schwerwiegendes unerwünschtes Ereignis; TSS: Gesamt-Symptomscore; UE: Unerwünschtes Ereignis.

## 2.4 Statistische Methoden

### Analysepopulationen der RCT-Phase über 24 Wochen (Teil 2):

- **Intention-to-Treat (ITT)-Population:** Alle randomisierten Personen, unabhängig davon, ob sie das Studienmedikament erhalten haben oder nicht; sie werden gemäß der randomisierten Behandlung analysiert und nicht nach der Dosis, die sie tatsächlich erhalten haben, falls diese abweicht.
- **Per-Protokoll (PP)-Population:** Alle Personen, die nach dem Zufallsprinzip einer Behandlung zugewiesen wurden,
  - mit einer bestätigten ISM-Diagnose gemäß WHO-Diagnosekriterien;
  - die keine wesentlichen Verstöße gegen folgende Einschlusskriterien aufweisen:
    - bestätigte Diagnose der systemischen Mastozytose und Prüfung von B- und C-Befunden,
    - mittelschwere bis schwere Symptome (TSS  $\geq$  28),
    - Kortikosteroid-Dosis:  $\leq$  20 mg/d Prednison oder Äquivalent;

- die keine wesentlichen Verstöße gegen folgende Ausschlusskriterien aufweisen:
  - Diagnose folgender Subklassen der systemischen Mastozytose: Kutane Mastozytose; Schwelende systemische Mastozytose; Systemische Mastozytose mit assoziierter hämatologischer Neoplasie; Aggressive systemische Mastozytose; Mastzelleukämie; Mastzellsarkom.
  - Erhalt eines der folgenden Medikamente, Therapien oder Verfahren: Tyrosinkinase-Inhibitor, antineoplastische Arzneimitteltherapie, Strahlentherapie oder Therapie mit PUVA (Psoralenen mit UV-A (langwelliges UV-Licht)), hämatopoetische Wachstumsfaktoren, größere chirurgische Eingriffe,
  - vorangegangene oder aktuelle klinisch bedeutsame Krankheit;
- die einen TSS-Baseline-Score  $\geq 28$  aufwiesen;
- die zu keinem Zeitpunkt ein falsches Studienmedikament erhalten haben (d. h. ein Placebo-Patient hat versehentlich Avapritinib erhalten oder umgekehrt); und
- die keine begleitende Behandlung mit Tyrosinkinase-Inhibitoren erhalten haben.

Die Personen in dieser Population werden entsprechend dem Behandlungsarm analysiert, dem sie zufällig zugewiesen wurden.

- Sicherheitspopulation: Alle Personen, die mindestens eine Dosis von Avapritinib oder Placebo erhalten haben. Die Sicherheitspopulation wird nach der tatsächlich erhaltenen Behandlung (Avapritinib oder Placebo) analysiert. Wenn eine Person zu irgendeinem Zeitpunkt das falsche Studienmedikament erhalten hat (d. h., eine Person des Placebo-Arms hat versehentlich Avapritinib erhalten oder umgekehrt), wird die Person dem Avapritinib-Arm zugeordnet.

Im SAP werden für die Effektivitätspunkte zwei Populationen beschrieben, die ITT- und die PP-Population. Beim primären Endpunkt und wichtigen sekundären Endpunkten wird im Studienbericht gemäß SAP die ITT-Population für die Hauptanalyse verwendet und die PP-Population als Sensitivitätsanalyse. Hingegen wurden in Modul 4 nur die Ergebnisse für die PP-Population dargestellt. Eine Begründung für die ausschließliche Darstellung der PP-Population in Modul 4 wurde vom pU nicht vorgelegt. Eine ergebnisgesteuerte Berichterstattung kann nicht ausgeschlossen werden.

Die beiden Populationen unterscheiden sich u. a. in Bezug auf die Erfüllung wichtiger Einschlusskriterien zu Baseline. Der im vorliegenden Fall wesentliche Unterschied zwischen der ITT- und der PP-Population liegt darin, dass bei der PP-Population alle Personen zu Baseline einen ISM-SAF TSS  $\geq 28$  aufweisen, wohingegen bei der ITT-Population alle randomisierten Personen einbezogen wurden. Diese Differenz konnte entstehen, da die Entscheidung zum Einschluss einer Person mehrere Wochen vor der Baseline-Erhebung getroffen wurde. Obwohl ein TSS  $\geq 28$  ein Einschlusskriterium darstellt (welches während der Screening-Periode für die betreffenden Personen auch erfüllt wurde), hatten von der ITT-Population 18 Personen (12,8 %) des Avapritinib-Arms und 4 (5,6 %) des Placebo-Arms zu Baseline einen TSS  $< 28$ . Änderungen bzw. Schwankungen des ISM-SAF TSS zwischen Screening-Phase und Baseline-Erhebung bedingen demzufolge die Unterschiede in der Anzahl an Personen in der ITT-Population im Vergleich zur PP-Population. Insgesamt umfasst die PP-Population knapp 90 % der ITT-Population, sodass gut 10 % der Studienpopulation aus den Analysen ausgeschlossen werden. Der Unterschied an ausgeschlossenen Personen zwischen den Armen ist größer als 5 %. Analysen innerhalb der PP-Population stellen einen Bruch zur Randomisierung dar und können daher mit Confounding einhergehen. Aus diesem Grund wird für die Nutzenbewertung die ITT-Population herangezogen, da sie alle randomisierten Personen umfasst und im SAP präspezifiziert war. Die Analysen der PP-Population werden im Anhang abgebildet.

## Datenschnitte

Im Rahmen der Zulassung wurden zur 24-wöchigen RCT-Phase (Teil 2) der Studie PIONEER Daten von zwei Datenschnitten der Studie PIONEER ausgewertet: der geplante Datenschnitt vom 23.06.2022 sowie ein von der European Medicines Agency (EMA) angeforderter Datenschnitt vom 07.04.2023. Zur Nutzenbewertung liegt ausschließlich der Datenschnitt vom 23.06.2022 der Studie PIONEER vor. Dieser beinhaltet die vollständig abgeschlossene 24-wöchige RCT-Phase (Teil 2) sowie Daten zur einarmigen Langzeitnachbeobachtung (Teil 3). Diese ist jedoch aufgrund der kurzen Beobachtungsdauer wenig aussagekräftig.

## Präspezifizierte Subgruppenanalysen

Für den primären Endpunkt, wichtige sekundäre Endpunkte (key secondary endpoints) sowie UE wurden Subgruppenanalysen a priori definiert und im Studienbericht abgebildet. Im Rahmen dieser Analysen wurden jedoch keine Interaktionstests durchgeführt. Die in Model 4 dargestellten Subgruppenanalysen enthalten Interaktionstests, allerdings beziehen sich die Berechnungen auf die PP-Population und werden daher nicht herangezogen.

## Präspezifizierung geplanter und durchgeführter Analysen

### ISM-SAF

Der primäre Endpunkt sowie (wichtige) sekundäre Endpunkte wurden mit dem ISM-SAF gemessen. Folgende Operationalisierungen wurden für den ISM-SAF im SAP beschrieben:

- Mittlere Veränderung des ISM-SAF zu Baseline für den Gesamtscore TSS (primärer Endpunkt), die Domänenscores GSS, SSS und NSS, die Items sowie das (schwerste) Leitsymptom und die/das (schwerste) Leitdomäne/-symptomcluster;
- Anteil an Personen mit einer Reduktion um  $\geq 30\%$  und  $\geq 50\%$  des TSS.

Bei der mittleren Veränderung ist anzumerken, dass für den Gesamtscore TSS ein anderes statistisches Analyseverfahren gewählt wurde als für die Domänenscores GSS, SSS und NSS, das (schwerste) Leitsymptom und die/das (schwerste) Leitdomäne/-symptomcluster. Für den TSS wurde eine Kovarianzanalyse (ANCOVA) berechnet, die für den Randomisierungs-Stratifizierungsfaktor (Serumtryptase) und den Baseline-ISM-Status (mittelschwer vs. schwer) adjustiert wurde. Dahingegen wurde die Signifikanz bei den anderen Scores mit einem Zwei-Stichprobent-Test berechnet, der nicht für den Randomisierungs-Stratifizierungsfaktor oder Baseline-ISM-Status adjustiert wurde.

Bei der prädefinierten Responderanalyse wurde als Responsekriterium eine Reduktion um mindestens 30 % und 50 % des TSS-Baseline-Werts festgelegt. Aufgrund des Einschlusskriteriums  $TSS \geq 28$  mussten die Personen mindestens eine Verbesserung um 8,4 bzw. 14 Punkte erzielen, um als Responder zu gelten. Beide Werte liegen unter der im IQWIG-Methodenpapier beschriebenen Responseschwelle von 15 % der Skalenspannbreite, diese beträgt 16,5 Punkte. Es ist unklar, wie hoch der Anteil an Personen war, deren Wert unter 16,5 Punkte lag, und die als Responder gezählt wurden; insbesondere da 18 Personen des Avapritinib-Arms (12,8 %) und 4 (5,6 %) des Placebo-Arms einen Baseline-Wert unter 28 aufwiesen.

In Modul 4 wurde außerdem eine Responderanalyse mit dem Responsekriterium „Verbesserung um  $\geq 15\%$  der Skalenspannweite“ des TSS sowie der Domänen/Cluster für die PP-Population dargestellt.

Aufgrund der genannten Unsicherheiten wird für die Nutzenbewertung die Mittlere Veränderung der ITT-Population gegenüber Baseline herangezogen. Darüber hinaus wird im Anhang die Responderanalyse zur Verbesserung um  $\geq 15\%$  der Skalenspannweite des TSS sowie der Domänen/Cluster der PP-Population abgebildet.

Weitere PRO

Im SAP wurde für die (weiteren/explorativen) patientenberichteten Endpunkte „EQ-5D-VAS“, „PGIS“, „PGIC“, „MC-QoL“ und „SF-12“ eine deskriptive Auswertung festgelegt (mittlere Veränderung gegenüber Baseline), die Berechnung von Effektschätzern war nicht vorhergesehen.

In Modul 4 des Dossiers wurden für die PP-Population neben Auswertungen zur mittleren Veränderung (inkl. Effektschätzer) auch Responderanalysen dargestellt.

- Für die PRO zur Morbidität (EQ-5D-VAS, PGIC, PGIS) wurde der Anteil an Personen mit einer mindestens 15%igen Verbesserung der Skalenspannweite des Instruments berechnet. Diese Responderanalysen entsprechen den Kriterien des IQWiG-Methodenpapiers.
- Für den SF-12 wurde in Modul 4 die mittlere Veränderung (inkl. Effektschätzer) und die Zeit bis zu einer Verbesserung um 15 % der Skalenspannweite dargestellt. Da keine relevant unterschiedlichen Beobachtungszeiten zwischen den Studienarmen vorliegen, stellt die Zeit bis zur ersten Veränderung keine sinnvolle Auswertung der Daten dar.

Für die Nutzenbewertung werden die für die ITT-Population durchgeführten Analysen herangezogen (mittlere Veränderung). Im Anhang werden für die PRO zur Morbidität Responderanalysen und für die PRO zur Lebensqualität die mittlere Veränderung zu Baseline inkl. Effektschätzer der PP-Population dargestellt.

**2.5 Verzerrungspotential auf Studien- und Endpunktebene**

*Tabelle 8: Verzerrungspotential der 24-wöchigen RCT-Phase (Teil 2) der Studie PIONEER*

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
PIONEER	Ja <sup>1)</sup>	Ja	Ja	Ja	Nein	Nein	Niedrig

<sup>1)</sup> Stratifizierte Randomisierung mittels Interactive Response Technology, Zuteilungsverhältnis 2:1.

Abkürzungen: RCT: Randomisierte kontrollierte Studie.

Für die Nutzenbewertung werden die Analysen zur ITT-Population herangezogen. Das Verzerrungspotential auf Studienebene wird als niedrig eingeschätzt.

Das Verzerrungspotential der für die Nutzenbewertung relevanten Endpunkte, für die Effektschätzer vorliegen, wird in Tabelle 9 dargestellt. Für die Endpunkte „Gesamtüberleben“ und „ISM-SAF TSS“ (Gesamtscore) wird das Verzerrungspotential als niedrig bewertet. Für die ISM-SAF-Domänenscores SSS, GSS, NSS, dem (schwersten) Leitsymptom und der/das (schwerste/n) Leitdomäne/-symptomcluster wird es als unklar bewertet. Für die weiteren Endpunkte liegen für die ITT-Population keine Effektschätzer vor.

Im Anhang wurden Analysen zur PP-Population (mit Effektschätzern) dargestellt. Diese wurden aus Modul 4 übernommen, in welchem nur Auswertungen zur PP-Population abgebildet waren. Hierbei kann eine ergebnisgesteuerte Berichterstattung nicht ausgeschlossen werden, da im SAP primär Auswertungen zur ITT-Population geplant waren. Bei der PP-Population liegt durch den nachträglichen Ausschluss von Personen(daten) ein Bruch der Randomisierung vor, sodass nicht

mehr von einer zufälligen Gruppeneinteilung ausgegangen werden kann. Im Avapritinib-Arm wurden 18 Personen (12,8 %) ausgeschlossen, im Placebo-Arm 4 Personen (5,4 %). Der Unterschied zwischen den Armen ist größer 5 %. Aus den genannten Gründen liegt für die im Anhang dargestellten Analysen zur PP-Population ein hohes Verzerrungspotential vor.

*Tabelle 9: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie PIONEER*

Studie PIONEER Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Gesamtüberleben	Ja	Ja <sup>1)</sup>	Nein	Nein	Niedrig
ISM-SAF: TSS <sup>2)</sup>	Ja	Ja <sup>3)</sup>	Nein	Nein <sup>3)</sup>	Niedrig
ISM-SAF: SSS, GSS, NSS, (schwerstes) Leitsymptom, (schwerste/s) Leitdomäne/-symptomcluster <sup>2)</sup>	Ja	Ja	Nein	Unklar <sup>4)</sup>	Unklar

<sup>1)</sup> Gesamtüberleben wurde im Rahmen der Beobachtung der Sicherheit als UE in der Sicherheitspopulation erhoben.

<sup>2)</sup> Die Bewertung gilt für die Analyse der mittleren Veränderung der ITT-Population.

<sup>3)</sup> Sensitivitätsanalysen zum primären Endpunkt bestätigen die Ergebnisse (MCMC-Imputation für fehlende Werte).

<sup>4)</sup> Bei der Analyse der Domänenscores wurden im Gegensatz zum Gesamtscore TSS die Randomisierungsstrata (Serumtryptase) und Baseline-ISM-Status (mittelschwer vs. schwer) nicht berücksichtigt. Dies war zwar präspezifiziert, ist methodisch jedoch nicht plausibel und nicht kongruent zum Gesamtscore.

Abkürzungen: GSS: Gastrointestinaler Symptomscore; ISM: Indolente systemische Mastozytose; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; ITT: Intention-to-Treat; MCMC: Markov Chain Monte Carlo; NSS: Neurokognitives Symptomcluster; SSS: Haut-Symptomscore; TSS: Gesamt-Symptomscore; UE: Unerwünschtes Ereignis.



### 3 Ergebnisse der eingeschlossenen Studie

Die folgenden Ergebnisse beziehen sich auf die 24-wöchige doppelblinde RCT-Phase (Teil 2) der Studie PIONEER.

#### 3.1 Studiencharakteristika und Studienmedikation

Tabelle 10: Allgemeine Angaben; Studie PIONEER

Studie PIONEER Allgemeine Angaben	Avapritinib + BSC N = 141	Placebo + BSC N = 71
ITT-Population <sup>1)</sup>	141 (100)	71 (100)
PP-Population	123 (87,2)	67 (94,4)
Sicherheitspopulation	141 (100)	71 (100)
Abbruch der Studie, n (%)	5 (3,5)	5 (7,0)
Aufgrund von:		
UE	1 (0,7)	0
Krankheitsprogression	1 (0,7)	0
Entscheidung Ärztin/Arzt	1 (0,7)	0
Lost to Follow-up	0	0
Tod	0	0
Protokollverletzung	0	0
Entzug der Einverständniserklärung	2 (1,4)	3 (4,2)
Anderes	0	2 (2,8)
Abbruch der Einnahme der Studienmedikation	141 (100)	71 (100)
Aufgrund von:		
UE	2 (1,4)	1 (1,4)
Anderes	139 (98,6)	70 (98,6)
Davon Übergang zur Langzeitbeobachtung (Teil 3) <sup>2)</sup>	135 (95,7)	66 (93,0)
Weitere Gründe	k. A.	k. A.
Personen, die die 24-wöchige RCT-Phase (Teil 2) abgeschlossen haben, n (%)	136 (96,5)	66 (93,0)
Sicherheits-Follow-up	k. A.	k. A.
Mediane Behandlungsdauer in Monaten (min; max)	5,55 (0,7; 6,5)	5,55 (0,8; 6,4)
Mediane Beobachtungsdauer in Monaten (min; max)	k. A.	k. A.

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

<sup>2)</sup> Keine Nachbeobachtung.

Abkürzungen: BSC: Best Supportive Care; ITT: Intention-to-Treat; k. A.: keine Angabe; PP: Per Protokoll; RCT: Randomisierte kontrollierte Studie; UE: Unerwünschtes Ereignis.

Tabelle 11: Charakterisierung der Studienpopulation; Studie PIONEER, ITT-Population

<b>Studie PIONEER Charakterisierung der Studienpopulation</b>	<b>Avapritinib 25 mg N = 141</b>	<b>Placebo N = 71</b>
<i>Alter (Jahre)</i> MW (SD) Median (min; max)	48,7 (11,7) 50 (18; 77)	52,2 (12,5) 54 (26; 79)
<i>Altersgruppe (Jahre), n (%)</i> < 65 Jahre ≥ 65 Jahre	132 (93,6) 9 (6,4)	60 (84,5) 11 (15,5)
<i>Geschlecht, n (%)</i> männlich weiblich	41 (29,1) 100 (70,9)	17 (23,9) 54 (76,1)
<i>Abstammung (genetisch), n (%)</i> weiß asiatisch unbekannt andere	109 (77,3) 1 (0,7) 27 (19,1) 4 (2,8)	61 (85,9) 0 8 (11,3) 2 (2,8)
<i>Region, n (%)</i> Nordamerika Europa Deutschland	60 (42,6) 81 (57,4) 19 (13,5)	33 (46,5) 38 (53,5) 15 (21,1)
<i>BMI zu Baseline</i> MW (SD) Median (min; max)	N = 137 28,3 (5,4) 27,8 (17,6; 42,0)	N = 70 29,5 (5,3) 29,0 (19,7; 43,7)
<i>Krankheitssymptomatik, ISM-SAF TSS zu Baseline<sup>1)</sup></i> MW (SD) Median (min; max)	N = 139 50,2 (19,2) 47,9 (12,1; 102,7)	52,4 (19,8) 47,8 (18,0; 104,4)
<i>Serumtryptase-Konzentration (ng/ml) zu Baseline</i> MW (SD) Median (min; max) < 20 ng/ml, n (%)	57,6 (54,4) 38,4 (3,6; 256,0) 28 (19,9)	67,6 (74,3) 43,7 (5,7; 501,6) 15 (21,1)
<i>KIT-D816V-Mutationslast im Blut zu Baseline, bestimmt anhand der MAF mittels ddPCR (zentraler Test)</i> MW (SD) Median (min; max)	2,57 (6,1) 0,39 (0,00; 41,29)	3,57 (7,6) 0,26 (0,00; 36,74)
<i>Mastzellen im Knochenmark zu Baseline (zentrale Pathologieuntersuchung) (Anteil in %)</i> MW (SD) Median (min; max)	11,0 (11,1) 7,0 (1,0; 50,0)	12,2 (12,6) 7,00 (1,0; 70,0)
<i>Mastzellen in der Haut (Anzahl/mm<sup>2</sup>) (zentraler Test) – läsional</i> MW (SD) Median (min; max)	N = 107 561,2 (399,7) 459 (89; 2.870)	N = 60 637,6 (488,7) 617 (83; 2.609)
<i>Mastzellen in der Haut (Anzahl/mm<sup>2</sup>) (zentraler Test) – nicht-läsional</i> MW (SD) Median (min; max)	N = 106 152,2 (76,3) 132 (35; 659)	N = 60 165,9 (166,0) 129 (59; 1337)
<i>Begleitende Therapie (BSC) zu Studienbeginn, n (%)</i> ja nein	140 (99,3) 1 (0,7)	71 (100) 0

<b>Studie PIONEER Charakterisierung der Studienpopulation</b>	<b>Avapritinib 25 mg N = 141</b>	<b>Placebo N = 71</b>
<i>Ort der SM-Beteiligung, n (%)</i>		
Knochenmark	136 (96,5)	69 (97,2)
Leber	3 (2,1)	3 (4,2)
Magen-Darm-Trakt	22 (15,6)	11 (15,5)
Lymphknoten	3 (2,1)	2 (2,8)
Milz	4 (2,8)	3 (4,2)
Haut	102 (72,3)	53 (74,6)
Andere	2 (1,4)	4 (5,6)
Mediane Zeit ab SM-Diagnose in Monaten (min; max)	60,5 (-2,0; 631,9)	71,6 (-2,1; 319,3)
<i>ECOG-PS, n (%)</i>		
0	68 (48,2)	33 (46,5)
1	63 (44,7)	34 (47,9)
2	10 (7,1)	4 (5,6)
3	0	0
<i>Medizinische Vorgeschichte von Anaphylaxie-Ereignissen, n (%)</i>		
ja	24 (17,0)	10 (14,1)
nein	117 (83,0)	61 (85,9)
<i>Häufigste (≥ 20 %) Vor- und Begleiterkrankungen mit Symptomatik nach SOC gemäß MedDRA</i>		
Erkrankungen des Gastrointestinaltrakts	114 (80,9)	55 (77,5)
Erkrankungen der Skelettmuskulatur und des Bindegewebes	112 (79,4)	58 (81,7)
Erkrankungen der Haut und des subkutanen Gewebes	110 (78,0)	52 (73,2)
Allgemeine (Begleit-)Erkrankungen	96 (68,1)	39 (54,9)
Erkrankungen des Nervensystems	95 (67,4)	44 (62,0)
Gefäßerkrankungen	85 (60,3)	48 (67,6)
Stoffwechsel- und Ernährungsstörungen	76 (53,9)	35 (49,3)
Psychiatrische Erkrankungen	69 (48,9)	30 (42,3)
Erkrankungen der Atemwege, des Brustraums u. Mediastinums	52 (36,9)	24 (33,8)
Erkrankungen des Immunsystems	46 (32,6)	24 (33,8)
Erkrankungen des hormonellen Systems	35 (24,8)	15 (21,1)

<sup>1)</sup> Der ISM-SAF TSS musste mind. 28 betragen, um in die Studie aufgenommen zu werden. Allerdings sind zu Baseline auch Werte unter 28 möglich, da die Messung zur Studieneignung mehrere Wochen vor Baseline durchgeführt wurde. Insgesamt hatten 18 Personen (12,8 %) des Avapritinib-Arms und 4 (5,6 %) des Placebo-Arms zu Baseline einen TSS < 28.

Abkürzungen: BSC: Best Supportive Care; ddPCR: Droplet Digital Polymerase Chain Reaction; ECOG-PS: Eastern Cooperative Oncology Group Performance Status; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; ITT: Intention-to-Treat; MAF: Mutationsallelfraction; MedDRA: Medical Dictionary for Regulatory Activities; MW: Mittelwert; SD: Standardabweichung; SM: Systemische Mastozytose; SOC: Systemorganklasse; TSS: Gesamt-Symptomscore.

### Protokollverletzungen

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

### Begleitmedikation

Anhand der in der Studie PIONEER dokumentierten Begleitmedikation ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens (keine Unterschiede zwischen den Armen ≥ 10 %). Am häufigsten wurden Antihistaminika zur systemischen Anwendung, Arzneimittel gegen Magengeschwüre und gastroösophageale Refluxkrankheit, andere systemische Arzneimittel gegen obstruktive Atemwegserkrankungen, intestinale entzündungshemmende Mittel und einfache Kortikoide zur systemischen Anwendung verordnet.

## 3.2 Mortalität

Während der 24-wöchigen RCT-Phase der Studie PIONEER (Teil 2) wurde kein Todesfall beobachtet.

## 3.3 Morbidität

Zur Bewertung der Morbidität werden die Analysen zur ITT-Population der 24-wöchigen RCT-Phase (Teil 2) herangezogen. Für die PRO „ISM-SAF“, „PGIS“ und „EQ-5D-VAS“ wird die a priori definierte mittlere Veränderung dargestellt. Im Anhang werden für die PRO die Responderanalysen inkl. Effektschätzer zur PP-Population dargestellt. Die Begründung für dieses Vorgehen wurde in Kapitel 2.4 beschrieben.

### 3.3.1 Symptome mittels „ISM-SAF“ und „PGIS“

#### ISM-SAF

Tabelle 12: Symptome mittels ISM-SAF; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022)

Studie PIONEER ISM-SAF	Avapritinib + BSC N = 141	Placebo + BSC N = 71
<b>Gesamt-Symptomscore (TSS)<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	139 (98,6)	71 (100)
MW (SD)	50,17 (19,15)	52,43 (19,82)
<i>Veränderung zu Woche 24</i>		
n (%)	128 (90,8)	65 (92,6)
LS Mean (SE) <sup>2)</sup>	-15,58 (1,54)	-9,15 (2,01)
LS Mean difference[95%-KI]; p-Wert <sup>2)</sup>	-6,43 [-10,90; -1,96]; 0,003	
Hedges' g	k. A.	
<b>Gastrointestinaler Symptomscore (GSS)<sup>3)</sup></b>		
<i>Baseline</i>		
n (%)	139 (98,6)	71 (100)
MW (SD)	9,71 (6,20)	10,41 (6,85)
<i>Veränderung zu Woche 24</i>		
n (%)	131 (92,9)	66 (93,0)
MW (SD)	-3,74 (4,87)	-3,14 (4,66)
MWD [95%-KI]; p-Wert <sup>4)</sup>	-0,60 [-2,03; 0,83]; 0,410	
<b>Haut-Symptomscore (SSS)<sup>3)</sup></b>		
<i>Baseline</i>		
n (%)	139 (98,6)	71 (100)
MW (SD)	15,50 (7,12)	17,20 (7,61)
<i>Veränderung zu Woche 24</i>		
n (%)	131 (92,9)	66 (93,0)
MW (SD)	-5,87 (6,49)	-2,64 (4,11)
MWD [95%-KI]; p-Wert <sup>4)</sup>	-3,23 [-4,73; -1,73]; < 0,0001	
Hedges' g	k. A.	

Studie PIONEER ISM-SAF	Avapritinib + BSC N = 141	Placebo + BSC N = 71
<b>Neurokognitives Symptomcluster (NSS)<sup>3)</sup> – ergänzende Darstellung</b>		
<i>Baseline</i>		
n (%)	139 (98,6)	71 (100)
MW (SD)	13,04 (6,78)	12,71 (6,97)
<i>Veränderung zu Woche 24</i>		
n (%)	131 (92,9)	66 (93,0)
MW (SD)	-4,03 (5,22)	-2,81 (4,29)
MWD [95%-KI]; p-Wert <sup>4)</sup>	-1,22 [-2,69; 0,25]; 0,102	
<b>(Schwerstes) Leitsymptom<sup>5)</sup></b>		
<i>Baseline</i>		
n (%)	139 (98,6)	71 (100)
MW (SD)	7,66 (1,69)	7,91 (1,68)
<i>Veränderung zu Woche 24</i>		
n (%)	131 (92,9)	66 (93,0)
MW (SD)	-2,22 (2,30)	-1,42 (1,88)
MWD [95%-KI]; p-Wert <sup>4)</sup>	-0,80 [-1,45; -0,16]; 0,015	
Hedges' g	k. A.	
<b>(Schwerste/s) Leitdomäne/-symptomcluster<sup>3)</sup></b>		
<i>Baseline</i>		
n (%)	139 (98,6)	71 (100)
MW (SD)	17,51 (6,03)	18,58 (6,40)
<i>Veränderung zu Woche 24</i>		
n (%)	131 (92,9)	66 (93,0)
MW (SD)	-6,52 (6,22)	-3,36 (4,49)
MWD [95%-KI]; p-Wert <sup>4)</sup>	-3,16 [-4,69; -1,63]; < 0,0001	
Hedges' g	k. A.	

<sup>1)</sup> Der Gesamtscore kann Werte zwischen 0 und 110 erreichen. Je höher die Punktezah, desto stärker ausgeprägt die Symptome.

<sup>2)</sup> ANCOVA-Modell, das für den Randomisierungs-Stratifizierungsfaktor (Serumtrypsinase) und den Baseline-ISM-Status (mittelschwer vs. schwer) kontrolliert wurde, einseitiger p-Wert.

<sup>3)</sup> Die Domänenscores können Werte zwischen 0 und 30 erreichen. Je höher die Punktezah, desto stärker ausgeprägt die Symptome.

<sup>4)</sup> MWD mit 95%-KI und Zwei-Stichproben-t-Test, es wurde nicht für den Randomisierungs-Stratifizierungsfaktor kontrolliert.

<sup>5)</sup> Die Items können Werte zwischen 0 und 10 erreichen. Je höher die Punktezah, desto stärker ausgeprägt die Symptome.

Abkürzungen: ANCOVA: Kovarianzanalyse; BSC: Best Supportive Care; ISM: Indolente systemische Mastozytose; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; MWD: Mittelwertdifferenz; SD: Standardabweichung; SE: Standardfehler.

**PGIS**

Tabelle 13: Symptome mittels PGIS; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022)

Studie PIONEER PGIS <sup>1)</sup>	Avapritinib + BSC N = 141	Placebo + BSC N = 71
<i>Baseline</i> n (%) MW (SD)	135 (95,7) 2,7 (0,89)	68 (95,8) 2,7 (0,92)
<i>Veränderung zu Woche 24</i> n (%) MW (SD)	121 (85,8) -0,8 (1,07)	60 (84,5) -0,4 (0,99)
MWD [95%-KI]; p-Wert	k. A.	

<sup>1)</sup> Werte zwischen 0 (keine Symptome) und 4 (Symptome lassen sich nicht ignorieren).

Abkürzungen: BSC: Best Supportive Care; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MW: Mittelwert; MWD: Mittelwertdifferenz; PGIS: Patient Global Impression of Severity; SD: Standardabweichung.

**3.3.2 Allgemeiner Gesundheitszustand mittels „EQ-5D-VAS“****EQ-5D-VAS**

Tabelle 14: Allgemeiner Gesundheitszustand mittels EQ-5D-VAS; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022)

Studie PIONEER EQ-5D-VAS <sup>1)</sup>	Avapritinib + BSC N = 141	Placebo + BSC N = 71
<i>Baseline</i> n (%) MW (SD)	135 (95,7) 57,2 (17,61)	67 (94,4) 55,9 (17,85)
<i>Veränderung zu Woche 24</i> n (%) MW (SD)	121 (85,8) 6,6 (18,84)	59 (83,1) 0,1 (13,58)
MWD [95%-KI]; p-Wert	k. A.	

<sup>1)</sup> Werte zwischen 0 (denkbar schlechtesten Gesundheitszustand) und 100 (denkbar besten Gesundheitszustand).

Abkürzungen: BSC: Best Supportive Care; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MW: Mittelwert; MWD: Mittelwertdifferenz; SD: Standardabweichung.

### 3.3.3 Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin

Zwischen der Einwilligung zur Studienteilnahme und der ersten Dosis des Studienmedikaments hatten 7 Personen (5,0 %) des Avapritinib-Arms und keine des Placebo-Arms eine Anaphylaxie-Episode, die eine Epinephrin-Behandlung erforderte. Während der Behandlungsphase hatten 4 Personen (2,8 %) des Avapritinib-Arms und 3 (4,2 %) des Placebo-Arms eine anaphylaktische Reaktion. Bei 2 Personen (1,4 %) des Avapritinib-Arms und 3 (4,2 %) des Placebo-Arms war eine Behandlung mit Epinephrin erforderlich.

*Tabelle 15: Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022)*

Studie PIONEER Anaphylaxie-Episode mit Epinephrin-Einsatz	Avapritinib + BSC N = 141	Placebo + BSC N = 71
n (%)	2 (1,4)	3 (4,2)
Effektschätzer		k. A.

Abkürzungen: BSC: Best Supportive Care; ITT: Intention-to-Treat; k. A.: keine Angabe.

### 3.4 Lebensqualität

Zur Bewertung der Lebensqualität werden die Analysen zur ITT-Population der 24-wöchigen RCT-Phase (Teil 2) herangezogen. Es wird die a priori definierte mittlere Veränderung dargestellt. Im Anhang wird die mittlere Veränderung der PP-Population dargestellt, da diese Effektschätzer umfasst. Die Begründung für dieses Vorgehen wurde in Kapitel 2.4 beschrieben.

#### Gesundheitsbezogene Lebensqualität mittels SF-12

*Tabelle 16: Gesundheitsbezogene Lebensqualität mittels SF-12; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022)*

Studie PIONEER SF-12	Avapritinib + BSC N = 141	Placebo + BSC N = 71
<b>Physical Component Summary (PCS)<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	134 (95,0)	67 (94,4)
MW (SD)	33,83 (10,79)	34,07 (10,70)
<i>Veränderung zu Woche 24</i>		
n (%)	120 (85,1)	59 (83,1)
MW (SD)	5,60 (8,43)	3,08 (7,01)
MWD [95%-KI]; p-Wert		k. A.
<b>Mental Component Summary (MCS)<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	134 (95,0)	67 (94,4)
MW (SD)	40,60 (10,51)	40,99 (10,67)
<i>Veränderung zu Woche 24</i>		
n (%)	120 (85,1)	59 (83,1)
MW (SD)	3,92 (9,27)	1,62 (9,03)
MWD [95%-KI]; p-Wert		k. A.

<sup>1)</sup> Werte zwischen 0 (denkbar schlechteste Lebensqualität) und 100 (denkbar beste Lebensqualität).

Abkürzungen: BSC: Best Supportive Care; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MW: Mittelwert; MWD: Mittelwertdifferenz; SD: Standardabweichung; SF-12: Short Form 12 Health Survey.

**Krankheitsspezifische Lebensqualität mittels MC-QoL**
**Tabelle 17: Krankheitsspezifische Lebensqualität mittels MC-QoL; Studie PIONEER, ITT-Population  
(Datenschnitt: 23.06.2022)**

<b>Studie PIONEER MC-QoL</b>	<b>Avapritinib + BSC N = 141</b>	<b>Placebo + BSC N = 71</b>
<b>Total Score</b>		
<i>Baseline</i>		
n (%)	135 (95,7)	68 (95,8)
MW (SD)	57,50 (16,02)	57,47 (17,20)
<i>Veränderung zu Woche 24</i>		
n (%)	121 (85,8)	60 (84,5)
MW (SD)	-19,20 (18,19)	-9,55 (14,93)
MWD [95%-KI]; p-Wert	k. A.	
<b>Symptome<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	135 (95,7)	68 (95,8)
MW (SD)	65,14 (16,07)	63,93 (17,55)
<i>Veränderung zu Woche 24</i>		
n (%)	121 (85,8)	60 (84,5)
MW (SD)	-18,82 (19,18)	-10,32 (17,40)
MWD [95%-KI]; p-Wert	k. A.	
<b>Soziales Leben / Funktionieren<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	135 (95,7)	68 (95,8)
MW (SD)	55,84 (21,02)	55,33 (22,82)
<i>Veränderung zu Woche 24</i>		
n (%)	121 (85,8)	60 (84,5)
MW (SD)	-20,09 (21,00)	-8,77 (18,82)
MWD [95%-KI]; p-Wert	k. A.	
<b>Emotionen<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	135 (95,7)	68 (95,8)
MW (SD)	49,51 (23,39)	48,98 (22,86)
<i>Veränderung zu Woche 24</i>		
n (%)	121 (85,8)	60 (84,5)
MW (SD)	-18,73 (21,44)	-9,89 (17,83)
MWD [95%-KI]; p-Wert	k. A.	
<b>Haut<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	135 (95,7)	68 (95,8)
MW (SD)	55,49 (22,33)	61,89 (22,69)
<i>Veränderung zu Woche 24</i>		
n (%)	121 (85,8)	60 (84,5)
MW (SD)	-18,60 (26,06)	-9,31 (18,80)
MWD [95%-KI]; p-Wert	k. A.	

<sup>1)</sup> Werte zwischen 0 (geringe Beeinträchtigung der Lebensqualität) bis 100 (starke Beeinträchtigung der Lebensqualität).

Abkürzungen: BSC: Best Supportive Care; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MC-QoL: Mastocytosis Quality of Life Questionnaire; MW: Mittelwert; MWD: Mittelwertdifferenz; SD: Standardabweichung.



### 3.5 Sicherheit

Im Folgenden werden die UE der Sicherheitspopulation der 24-wöchigen RCT-Phase (Teil 2) zum Datenschnitt vom 23.06.2022 dargestellt.

*Tabelle 18: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022)*

<b>Studie PIONEER Zusammenfassung der UE<sup>1)</sup> Personen mit mindestens einem ...</b>	<b>Avapritinib + BSC N = 141 n (%)</b>	<b>Placebo + BSC N = 71 n (%)</b>	<b>Effektschätzer [95%-KI] p-Wert</b>
UE (ergänzend dargestellt)	128 (90,8)	66 (93,0)	k. A.
UE CTCAE-Grad ≥ 3	30 (21,3)	15 (21,1)	k. A.
SUE	7 (5,0)	8 (11,3)	k. A.
UE, das zum Abbruch der Studienmedikation führte	3 (2,1)	1 (1,4)	k. A.

<sup>1)</sup> Der pU legt keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Er beschreibt in Modul 4 zur PP-Population, dass im Rahmen der ISM keine erkrankungsbezogenen Ereignisse aufgetreten sind. Allerdings werden (erkrankungsbedingte) anaphylaktische Episoden ebenfalls als UE gewertet.

Abkürzungen: BSC: Best Supportive Care; CTCAE: Common Terminology Criteria for Adverse Events; ISM: Indolente systemische Mastozytose; k. A.: keine Angabe; KI: Konfidenzintervall; PP: Per Protokoll; pU: pharmazeutischer Unternehmer; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

### Unerwünschte Ereignisse

*Tabelle 19: UE mit Inzidenz ≥ 10 %; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022)*

<b>Studie PIONEER UE mit Inzidenz ≥ 10 % MedDRA-Systemorganklasse Preferred Term</b>	<b>Avapritinib + BSC N = 141 n (%)</b>	<b>Placebo + BSC N = 71 n (%)</b>	<b>Effektschätzer [95%-KI] p-Wert</b>
<b>Augenerkrankungen</b>	<b>28 (19,9)</b>	<b>10 (14,1)</b>	<b>k. A.</b>
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>52 (36,9)</b>	<b>27 (38,0)</b>	<b>k. A.</b>
Übelkeit	18 (12,8)	12 (16,9)	k. A.
Diarrhö	15 (10,6)	8 (11,3)	k. A.
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>51 (36,2)</b>	<b>15 (21,1)</b>	<b>k. A.</b>
<b>Infektionen und parasitäre Erkrankungen</b>	<b>49 (34,8)</b>	<b>22 (31,0)</b>	<b>k. A.</b>
COVID-19	17 (12,1)	8 (11,3)	k. A.
<b>Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen</b>	<b>14 (9,9)</b>	<b>10 (14,1)</b>	<b>k. A.</b>
<b>Untersuchungen</b>	<b>33 (23,4)</b>	<b>14 (19,7)</b>	<b>k. A.</b>
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>16 (11,3)</b>	<b>5 (7,0)</b>	<b>k. A.</b>
<b>Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen</b>	<b>32 (22,7)</b>	<b>19 (26,8)</b>	<b>k. A.</b>
<b>Erkrankungen des Nervensystems</b>	<b>54 (38,3)</b>	<b>24 (33,8)</b>	<b>k. A.</b>
Kopfschmerzen	27 (19,1)	14 (19,7)	k. A.
Schwindelgefühl	16 (11,3)	6 (8,5)	k. A.

Studie PIONEER UE mit Inzidenz $\geq$ 10 % MedDRA-Systemorganklasse Preferred Term	Avapritinib + BSC N = 141 n (%)	Placebo + BSC N = 71 n (%)	Effektschätzer [95%-KI] p-Wert
Psychiatrische Erkrankungen	19 (13,5)	5 (7,0)	k. A.
Erkrankungen der Atemwege, des Brustraums und Mediastinums	18 (12,8)	8 (11,3)	k. A.
Erkrankungen der Haut und des Unterhautzellgewebes	43 (30,5)	21 (29,6)	k. A.
Gefäßerkrankungen	28 (19,9)	9 (12,7)	k. A.
Flushing	13 (9,2)	3 (4,2)	k. A.

Abkürzungen: BSC: Best Supportive Care; MedDRA: Medical Dictionary for Regulatory Activities; k. A.: keine Angabe; KI: Konfidenzintervall; UE: Unerwünschtes Ereignis.

In der Studie PIONEER traten SUE bei 7 Personen (5 %) des Avapritinib-Arms und bei 8 Personen (11,3 %) des Placebo-Arms auf. In keinem Arm traten SUE auf Ebene der Systemorganklassen bei mehr als 5 % der Studienteilnehmenden auf.

Schwere UE (CTCAE-Grad 3 oder höher) traten bei 30 Personen (21,3 %) des Avapritinib-Arms und bei 15 Personen (21,1 %) des Placebo-Arms auf. In den Systemorganklassen „Erkrankungen des Immunsystems“ sowie „Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen“ traten bei  $\geq$  5 % der Studienteilnehmenden Ereignisse auf.

*Tabelle 20: Schwere UE (CTCAE-Grad 3 oder höher), die bei  $\geq$  5 % der Patientinnen und Patienten in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022)*

Studie PIONEER Schwere UE (CTCAE-Grad 3 oder höher) MedDRA-Systemorganklasse	Avapritinib + BSC N = 141 n (%)	Placebo + BSC N = 71 n (%)
Erkrankungen des Immunsystems	4 (2,8)	4 (5,6)
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	2 (1,4)	4 (5,6)

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; MedDRA: Medical Dictionary for Regulatory Activities; UE: Unerwünschtes Ereignis.

### Unerwünschte Ereignisse von besonderem Interesse

UESI traten bei 4 Personen (2,8 %) des Avapritinib-Arms und bei 3 Personen (4,2 %) des Placebo-Arms auf. In Tabelle 21 sind nur die Preferred Terms dargestellt, in denen Ereignisse aufgetreten sind.

Ein schwerwiegendes und schweres UESI (CTCAE-Grad 3 oder höher) trat bei einer Person des Placebo-Arms (1,4 %) im Preferred Term „Gemütszustand verändert“ (Kognitive Effekte) auf. Im Avapritinib-Arm wurden keine schwerwiegenden und schweren UESI verzeichnet.

*Tabelle 21: UESI nach Systemorganklasse und Preferred Term; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022)*

<b>Studie PIONEER UESI MedDRA-Systemorganklasse Preferred Term</b>	<b>Avapritinib + BSC N = 141 n (%)</b>	<b>Placebo + BSC N = 71 n (%)</b>
<b>Kognitive Effekte</b>	<b>4 (2,8)</b>	<b>3 (4,2)</b>
Erinnerungsvermögen eingeschränkt	2 (1,4)	1 (1,4)
Amnesie	1 (0,7)	0
Stimmungsänderung	1 (0,7)	0
Geistige Beeinträchtigung	0	1 (1,4)
Gemütszustand verändert	0	1 (1,4)
<b>Intrakranielle Blutungen</b>	<b>0</b>	<b>0</b>

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; UESI: Unerwünschtes Ereignis von besonderem Interesse.

## Ödeme

Ödeme traten bei 36 Personen (25,5 %) des Avapritinib-Arms und bei 8 Personen (11,3 %) des Placebo-Arms auf (siehe Tabelle 22). Es traten keine schwerwiegenden und schweren Ödeme (CTCAE-Grad 3 oder höher) auf.

*Tabelle 22: Ödeme als UESI nach Customized MedDRA Queries; Studie PIONEER, Sicherheitspopulation (Datenschnitt: 23.06.2022)*

<b>Studie PIONEER Ödeme Customized MedDRA Queries Preferred Term</b>	<b>Avapritinib + BSC N = 141 n (%)</b>	<b>Placebo + BSC N = 71 n (%)</b>
<b>Ödeme</b>	<b>36 (25,5)</b>	<b>8 (11,3)</b>
Ödem peripher	12 (8,5)	3 (4,2)
Gesichtsödem	10 (7,1)	1 (1,4)
Periorbitalödem	9 (6,4)	2 (2,8)
Augenlidödem	5 (3,5)	1 (1,4)
Flüssigkeitsretention	2 (1,4)	0
Schwellung des Augenlids	2 (1,4)	0
Augenödem	1 (0,7)	1 (1,4)
Generalisiertes Ödem	1 (0,7)	0
Ödem	1 (0,7)	0
Schwellung des Auges	0	1 (1,4)

Abkürzungen: CMQ: Customized MedDRA Queries; MedDRA: Medical Dictionary for Regulatory Activities; UESI: Unerwünschtes Ereignis von besonderem Interesse.

## 4 Diskussion der Methodik und Ergebnisse

### 4.1 Zulassungsstatus und Zulassungspopulation von Avapritinib

Die vorliegende Bewertung bezieht sich auf den Nutzen von Avapritinib bei erwachsenen Patientinnen und Patienten mit indolenter systemischer Mastozytose (ISM) mit mittelschweren bis schweren Symptomen, bei denen mit einer symptomatischen Behandlung keine ausreichende Kontrolle erzielt werden kann. Es ist die zweite Erweiterung des Anwendungsgebiets, Avapritinib ist bereits für folgende Anwendungsgebiete zugelassen:

- Inoperable oder metastasierte gastrointestinale Stromatumoren (GIST), die eine Thrombozyten-Wachstumsfaktor-Rezeptor-alpha (PDGFRA)-D842V-Mutation aufweisen.
- Aggressive systemische Mastozytose, systemische Mastozytose mit assoziierter hämatologischer Neoplasie oder Mastzellleukämie nach zumindest einer systemischen Therapie.

Für Patientinnen und Patienten mit ISM beträgt die empfohlene Dosis 25 mg, die gemäß Fachinformation oral eingenommen werden soll. Die Behandlung von ISM soll fortgesetzt werden, bis die Erkrankung fortschreitet oder eine inakzeptable Toxizität auftritt.

Für die Nutzenbewertung liegen Daten der pivotalen Studie PIONEER vor, in die erwachsene Patientinnen und Patienten mit bestätigter ISM-Diagnose eingeschlossen wurden, die trotz optimal eingestellter, stabiler symptomatischer Therapie (BSC) der systemischen Mastozytose mittelschwere bis schwere Symptome aufwiesen. Der Schweregrad wurde mit dem Gesamtscore des ISM-SAF bestimmt (Einschlusskriterium für Studie PIONEER TSS  $\geq$  28).

Die Dosierung von Avapritinib stimmt im für die Nutzenbewertung relevanten Teil 2 der Studie PIONEER mit der empfohlenen Dosis gemäß Fachinformation überein. Während der Behandlungsphase sollte die BSC-Therapie fortgesetzt werden. Eine Reduktion der BSC-Therapie war auf Empfehlung des ärztlichen Personals möglich. Hingegen geht aus der Fachinformation nicht eindeutig hervor, ob und in welcher Weise die BSC-Therapie fortgesetzt werden soll. Kortikosteroide sollten während der Behandlungsphase nur verabreicht werden, wenn die Person auch zu Beginn Kortikosteroide benötigte. Außerdem sollte die Dosis nicht erhöht werden. Es ist unklar, warum es diese Einschränkungen zu Kortikosteroiden gibt, da die Patientinnen und Patienten die bestmögliche BSC-Therapie während der Studie erhalten sollten.

An der Studie PIONEER haben 23 Studienzentren in Europa (inkl. Großbritannien) und 19 in Nordamerika teilgenommen. In den 5 Studienzentren in Deutschland wurden insgesamt 34 Personen (16,0 %) behandelt. Es wird davon ausgegangen, dass die Ergebnisse auf den deutschen Versorgungskontext übertragbar sind.

### 4.2 Design und Methodik der Studie

#### Zusammenfassung des Studiendesigns

Bei der Studie PIONEER handelt es sich um eine dreiteilige Phase-II-Studie zur Dosisfindung (Teil 1), zur Untersuchung der Wirksamkeit und Sicherheit von Avapritinib in Kombination mit BSC im Vergleich zu Placebo in Kombination mit BSC (Teil 2) sowie zur Untersuchung der Langzeitsicherheit von Avapritinib (Teil 3). Teil 1 der Studie war eine randomisierte 4-armige Dosisfindungsphase über mindestens 12 Wochen mit 39 Personen. Teil 2 war eine doppelblinde randomisiert-kontrollierte Studienphase (Avapritinib + BSC vs. Placebo + BSC) über mindestens 24 Wochen mit 212 Personen. Teil 3 beinhaltet eine einarmige, offene Langzeitbeobachtung, in die Patientinnen und Patienten nach Abschluss von Teil 1 und 2 übergehen konnten und die maximal 5 Jahre andauern soll.

Zur Nutzenbewertung wird die 24-wöchige doppelblinde randomisiert-kontrollierte doppelblinde Studienphase (Teil 2) der Studie PIONEER mit 212 Patientinnen und Patienten herangezogen. Die

Dosierung in Teil 2 erfolgte in der in der Fachinformation empfohlenen Dosis. Die doppelblinde RCT-Phase zur Dosisfindung (Teil 1) wird nicht herangezogen, da die geplante Behandlungszeit mit mindestens 12 Wochen deutlich kürzer ist als die von Teil 2 (mindestens 24 Wochen) und für die relevanten Effektivitätseindpunkte nur Analysen zu Zyklus 4 Tag 1 (12 Wochen) vorliegen. Darüber hinaus ist auch die Anzahl der eingeschlossenen Personen deutlich geringer als bei Teil 2 (Avapritinib 25 mg / Placebo: Teil 1: 10/9; Teil 2: 141/71). Auch die Langzeitbeobachtung (Teil 3) wird nicht herangezogen, da der vom pU zur Nutzenbewertung vorgelegte Datenschnitt aufgrund der kurzen Beobachtungszeit (Median: 5,16 Monate (MW: 6,78)) keine Aussagen zur Langzeitsicherheit und -wirksamkeit zulässt. Der spätere, im European Public Assessment Report (EPAR), dargestellte Datenschnitt vom 07.04.2023, bei dem die mittlere Beobachtungszeit für Teil 3 bei 10,4 Monaten (SD: 7,17) lag, wurde im Dossier nicht vorgelegt.

In der 24-wöchigen RCT-Phase (Teil 2) wurden die Patientinnen und Patienten im Verhältnis 2:1 in den Avapritinib- und Placebo-Arm randomisiert, stratifiziert nach dem Serumtrypsin-Level (< 20 ng/ml vs. ≥ 20 ng/ml). Maximal 20 % der Teilnehmenden sollten einen Serumtrypsin-Wert unter 20 ng/ml aufweisen. Die Behandlungsphase mit Avapritinib + BSC bzw. Placebo + BSC sollte mindestens 24 Wochen (6 Zyklen à 28 Tage) andauern. Folgende Kriterien führten zum Abbruch der Behandlung: Entwicklung inakzeptabler Toxizität (Intrakranielle Blutungen beliebigen Grades; anhaltende Neutropenie; Anämie oder Thrombozytopenie des Grades 3 oder 4), klinisches Fortschreiten der systemischen Mastozytose, Beginn einer neuen Therapie, Tod, Verlust der Nachbeobachtung, Protokollabweichung, Rücknahme der Einwilligungserklärung, Schwangerschaft oder Stillzeit, Entscheidung des Prüfpersonals, Non-Compliance, Entscheidung des Sponsors. Klinisches Fortschreiten (Progression) war definiert als Entwicklung eines assoziierten hämatologischen Neoplasmas, Fortschreiten zur Mastzelleukämie oder zur aggressiven systemischen Mastozytose mit neuen WHO C-Befunden, die auf systemische Mastozytose zurückzuführen sind. Der primäre Endpunkt zu den Symptomen der ISM wurde mittels ISM-SAF erhoben. Auf Anraten der U. S. Food and Drug Administration (FDA) wurde die statistische Auswertung des primären Endpunkts von "Anteil der Personen mit einer ≥ 30%igen Reduktion im TSS" in "Mittlere Veränderung des ISM-SAF TSS" im Laufe der Studie geändert (siehe Kapitel 2.3).

### **Studienpopulation**

Die Studienpopulation der 24-wöchigen doppelblinden RCT-Phase (Teil 2) bestand aus 212 Personen mit ISM, von denen 141 Personen in den Avapritinib-Arm und 71 in den Placebo-Arm randomisiert wurden. Es liegen keine Angaben vor, wie viele Personen für diesen Teil gescreent wurden.

Die Baseline-Charakteristika waren weitestgehend ausbalanciert zwischen den Behandlungsarmen. Das mediane Alter der Patientinnen und Patienten betrug 54 Jahre im Avapritinib-Arm und 50 Jahre im Placebo-Arm. Knapp drei Viertel der Teilnehmenden waren weiblich (72,6 %). Die Patientinnen und Patienten hatten einen medianen BMI von 27,8 bzw. 29,0 kg/m<sup>2</sup>. Die Mastozytose trat bei fast allen Patientinnen und Patienten im Knochenmark (96,7 %) auf, daneben waren häufig die Haut (73,1 %) und der Magen-Darm-Trakt (15,6 %) betroffen. Der mediane Serumtrypsin-Wert lag zu Baseline bei 38,4 ng/ml im Avapritinib-Arm und bei 43,7 ng/ml im Placebo-Arm. Nahezu alle Patientinnen und Patienten (99,5 %) erhielten zu Studienbeginn eine begleitende BSC-Therapie. Mehr als die Hälfte der Teilnehmenden (> 50 %) hatte zu Baseline Erkrankungen des Gastrointestinaltrakts, der Skelettmuskulatur und des Bindegewebes, der Haut und des subkutanen Gewebes, allgemeine (Begleit-)Erkrankungen, Erkrankungen des Nervensystems, Gefäßerkrankungen sowie Stoffwechsel- und Ernährungsstörungen.

Alle randomisierten Studienteilnehmenden haben mindestens eine Dosis der Studienmedikation erhalten. Damit entspricht die Sicherheitspopulation allen randomisierten Personen (ITT-Population).

Entgegen der Präspezifizierungen im SAP wurden in Modul 4 Analysen zur PP-Population dargestellt. Die beiden Populationen unterscheiden sich in Bezug auf das Einschlusskriterium ISM-SAF TSS  $\geq 28$ . Für alle Personen der ITT-Population lag der ISM-SAF TSS zum Zeitpunkt des Einschlusses in die Studie  $\geq 28$ . Für alle Personen der PP-Population lag der ISM-SAF TSS sowohl zum Zeitpunkt des Einschlusses in die Studie als auch zum Zeitpunkt der Baseline-Messung vor Behandlungsbeginn bei  $\geq 28$ . Da die Messung zum Studieneinschluss mehrere Wochen vor Baseline durchgeführt wurde, waren Schwankungen des ISM-SAF TSS möglich. 18 Personen (12,8 %) des Avapritinib-Arms und 4 Personen (5,6 %) des Placebo-Arms wiesen zu Baseline einen ISM-SAF TSS unter 28 auf. Diese Personen wurden bei der PP-Population ausgeschlossen, die laut SAP für Sensitivitätsanalysen des primären Endpunkts und wichtiger sekundärer Endpunkte gedacht war. Durch den Ausschluss dieser Personen wird ein Randomisierungsbruch angenommen. Der Unterschied an ausgeschlossenen Personen ist zwischen den Armen größer als 5 %. Für die Nutzenbewertung werden daher die prädefinierten Analysen zur ITT-Population herangezogen. Ergebnisse der PP-Population sind im Anhang dargestellt.

Die mediane Behandlungsdauer war mit 5,55 Monaten in beiden Studienarmen gleich.

### **Bewertung des Verzerrungspotentials und Aussagekraft der Studienergebnisse**

Das Verzerrungspotential der 24-wöchigen doppelblinden RCT-Phase der Studie PIONEER (Teil 2) wird auf Studienebene für die ITT-Population als niedrig bewertet.

## **4.3 Mortalität**

In der 24-wöchigen RCT-Phase (Teil 2) der Studie PIONEER traten keine Todesfälle auf. Auf Basis der 24-wöchigen Studienphase lassen sich keine Unterschiede in der Mortalität ableiten.

## **4.4 Morbidität**

Für die Endpunktkategorie „Morbidität“ legte der pU Analysen zu Symptomen, dem allgemeinen Gesundheitszustand, dem Anteil der Mastzellen im Knochenmark, dem Serumtryptase-Wert, der KIT-D816V-Mutationslast, der Mastrozytose auf der Haut sowie dem Gebrauch von BSC-Maßnahmen und Epinephrin vor.

Die detaillierte Bewertung der Endpunkte ist in Kapitel 2.3.2 beschrieben. Zusammenfassend wurden für die Nutzenbewertung Ergebnisse zu Symptomen, dem allgemeinen Gesundheitszustand und dem Einsatz von Epinephrin berücksichtigt. Abweichend von der Einschätzung des pU werden die laborparametrischen Analysen und die Auswertungen zur Mastrozytose auf der Haut aufgrund nicht gegebener bzw. unklarer Patientenrelevanz nicht dargestellt.

### **Symptome mittels „ISM-SAF“ und „PGIS“**

#### ISM-SAF

Für den ISM-SAF-Gesamtscore (TSS) wurde der Unterschied der mittleren Veränderung der beiden Behandlungsarme mittels ANCOVA ausgewertet, kontrolliert für den Randomisierungs-Stratifizierungsfaktor (Serumtryptase) und den Baseline-ISM-Status (mittelschwer vs. schwer). Es liegt ein statistisch signifikanter Unterschied zugunsten von Avapritinib gegenüber Placebo vor (LS Mean difference: -6,43 (95%-KI: [-10,90, -1,96]); einseitiger p-Wert = 0,003). Angaben zu Hedges' g zur Einschätzung der klinischen Relevanz liegen für die ITT-Population nicht vor.

Bei den Domänenscores und dem (schwersten) Leitsymptom bzw. der/dem (schwersten) Leitdomäne/-symptomcluster wurde die Signifikanz der Mittelwertdifferenz (MWD) mit einem Zwei-Stichproben-t-Test analysiert. Im Gegensatz zur Auswertung des Gesamtscores TSS wurde

nicht für den Randomisierungs-Stratifizierungsfaktor (Serumtryptase) und den Baseline-ISM-Status (mittelschwer vs. schwer) kontrolliert. Für die Domäne „Haut-Symptomscore“ (SSS) liegt ein statistisch signifikanter Unterschied zugunsten von Avapritinib gegenüber Placebo vor (MWD: -3,23 (95%-KI: [-4,73; -1,73]);  $p < 0,0001$ ). Für die Domäne „Gastrointestinaler Symptomscore“ (GSS) liegt kein statistisch signifikanter Unterschied zwischen Avapritinib und Placebo vor. Für das (schwerste) Leitsymptom liegt ein statistisch signifikanter Unterschied zugunsten von Avapritinib gegenüber Placebo vor (MWD: -0,80 (95%-KI: [-1,45; -0,16]);  $p = 0,015$ ). Für die/das (schwerste) Leitdomäne/-symptomcluster liegt ebenfalls ein statistisch signifikanter Unterschied zugunsten von Avapritinib gegenüber Placebo vor (MWD: -3,16 (95%-KI: [-4,69, -1,63]);  $p < 0,0001$ ).

### PGIS

Die Auswertungen zum PGIS erfolgten für die ITT-Population lediglich deskriptiv, sodass sie keine für die Nutzenbewertung relevante Aussage ermöglichen.

### **Allgemeiner Gesundheitszustand mittels „EQ-5D-VAS“**

Die Auswertungen des explorativ erhobenen Endpunkts „EQ-5D-VAS“ waren für die ITT-Population deskriptiv, sodass sie keine für die Nutzenbewertung relevanten Aussagen ermöglichen.

### **Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin**

Der Einsatz von Epinephrin zur Behandlung anaphylaktischer Episoden wurde deskriptiv dargestellt, Effektschätzer wurden nicht berechnet. Daher können keine für die Nutzenbewertung relevanten Aussagen abgeleitet werden.

Insgesamt zeigt sich für die Endpunktkategorie „Morbidity“ ein Vorteil von Avapritinib gegenüber Placebo im Bereich der Symptome, dessen klinische Relevanz aufgrund fehlender Werte für Hedges'  $g$  jedoch unklar ist.

## **4.5 Lebensqualität**

Für die Endpunktkategorie „Lebensqualität“ legte der pU Daten zur gesundheitsbezogenen Lebensqualität mittels SF-12 und der krankheitsspezifischen Lebensqualität mittels MC-QoL vor. Die Auswertungen der beiden explorativ erhobenen Endpunkte waren für die ITT-Population deskriptiv, sodass sie keine für die Nutzenbewertung relevanten Aussagen ermöglichen.

## **4.6 Sicherheit**

Die mediane Behandlungsdauer betrug in beiden Armen 5,55 Monate. Die UE wurden in dieser Zeit kontinuierlich erfasst, jedoch erfolgten die Abfragen mit Verabreichung der Studienmedikation zu den Studienvisiten. Bei der Interpretation der UE ist zu beachten, dass die erhobenen und in der Nutzenbewertung dargestellten UE möglicherweise Symptome der Grunderkrankung umfassen. Der pU beschreibt in Modul 4 zur PP-Population, dass keine erkrankungsbezogenen Ereignisse aufgetreten sind.

Es liegen keine Effektschätzer für die UE der ITT-Population vor. Von den UE jeglichen Schweregrads traten in der Systemorganklasse „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ deutlich mehr UE (Unterschiede  $> 15\%$ ) im Avapritinib-Arm als im Placebo-Arm auf (51 (36,2 %) vs. 15 (21,1 %)). Auch kamen Ödeme, die vom pU als häufig auftretende UE bei fortgeschrittener systemischer Mastozytose identifiziert wurden, im Avapritinib-Arm häufiger vor als im Placebo-Arm (36 (25,5 %) vs. 8 (11,3 %)).

Aus dem Sicherheitsbericht des EPAR gehen folgende Hinweise hervor: Die EMA stuft das Sicherheitsprofil von Avapritinib auf Basis der pivotalen Studien insgesamt als gut verträglich ein [6]. Es bestehen Unsicherheiten hinsichtlich der möglichen Auswirkungen auf den Krankheitsverlauf nach Absetzen von Avapritinib. Die Unklarheit hinsichtlich des karzinogenen Potentials von Avapritinib wird ebenfalls als bedenklich angesehen.



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

Die Angaben der Fachinformation sind zu berücksichtigen.

## 6 Zusammenfassung der Nutzenbewertung

Avapritinib ist neu zugelassen zur Behandlung erwachsener Patienten mit indolenter systemischer Mastozytose (ISM) mit mittelschweren bis schweren Symptomen, bei denen mit einer symptomatischen Behandlung keine ausreichende Kontrolle erzielt werden kann. Die Nutzenbewertung von Avapritinib basiert auf der zulassungsbegründenden Studie PIONEER. Bei der Studie PIONEER handelt es sich um eine dreiteilige Phase-II-Studie zur Dosisfindung (Teil 1, RCT, mind. 12 Wochen), zur Untersuchung der Wirksamkeit und Sicherheit von Avapritinib in Kombination mit BSC im Vergleich zu Placebo in Kombination mit BSC (Teil 2, RCT, mind. 24 Wochen) sowie zur Bestimmung der Langzeitwirksamkeit und -sicherheit (Teil 3, einarmig). Teilgenommen haben erwachsene Patientinnen und Patienten mit bestätigter ISM-Diagnose, die trotz optimal eingestellter stabiler symptomatischer Therapie der systemischen Mastozytose mittelschwere bis schwere Symptome aufwiesen (ISM-SAF TSS  $\geq 28$ ). Zur Nutzenbewertung wird die 24-wöchige RCT-Phase (Teil 2) der Studie PIONEER herangezogen.

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 Avapritinib 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 23: Zusammenfassende Darstellung der Ergebnisse der Studie PIONEER

Studie PIONEER Darstellung der Ergebnisse <sup>1)</sup>	Avapritinib + BSC N = 141 <sup>2)</sup>				Placebo + BSC N = 71 <sup>2)</sup>				Wirkstoff vs. Komparator	Effekt	
<b>Mortalität</b>											
<b>Gesamtüberleben</b>	N <sup>3)</sup>	Personen mit Ereignis, n (%)			N <sup>3)</sup>	Personen mit Ereignis, n (%)			Effektschätzer		
Todesfälle	141	0			71	0			-	↔	
<b>Morbidität</b>											
<b>Symptome mittels ISM-SAF</b>	Baseline		Änderung Woche 24 zu Baseline		Baseline		Änderung Woche 24 zu Baseline		Effektschätzer		
	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	LS Mean (SE) <sup>4)</sup>	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	LS Mean (SE) <sup>4)</sup>	LS Mean difference [95%-KI]; p-Wert <sup>4)</sup>		
Gesamt-Symptomscore (TSS) <sup>5)</sup>	139	50,17 (19,15)	128	-15,58 (1,54)	71	52,43 (19,82)	65	-9,15 (2,01)	-6,43 [-10,90; -1,96]; 0,003		↑↑ <sup>6)</sup>
	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	MWD [95%-KI]; p-Wert <sup>7)8)</sup>		
Gastrointestinaler Symptomscore <sup>9)</sup>	139	9,71 (6,20)	131	-3,74 (4,87)	71	10,41 (6,85)	66	-3,14 (4,66)	-0,60 [-2,03; 0,83]; 0,410		↔
Haut-Symptomscore <sup>9)</sup>	139	15,50 (7,12)	131	-5,87 (6,49)	71	17,20 (7,61)	66	-2,64 (4,11)	-3,23 [-4,73; -1,73]; < 0,0001		↑ <sup>6)</sup>
(Schwerstes) Leitsymptom <sup>10)</sup>	139	7,66 (1,69)	131	-2,22 (2,30)	71	7,91 (1,68)	66	-1,42 (1,88)	-0,80 [-1,45; -0,16]; 0,015		↑ <sup>6)</sup>
(Schwerste/s) Leitdomäne/-symptomcluster <sup>9)</sup>	139	17,51 (6,03)	131	-6,52 (6,22)	71	18,58 (6,40)	66	-3,36 (4,49)	-3,16 [-4,69; -1,63]; < 0,0001		↑ <sup>6)</sup>
<b>Symptome mittels PGIS<sup>11)</sup></b>	Baseline		Änderung Woche 24 zu Baseline		Baseline		Änderung Woche 24 zu Baseline		Effektschätzer		
	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)			
	135	2,7 (0,89)	121	-0,8 (1,07)	68	2,7 (0,92)	60	-0,4 (0,99)	k. A.		n. b.
<b>Allgemeiner Gesundheitszustand mittels EQ-5D-VAS<sup>12)</sup></b>	Baseline		Änderung Woche 24 zu Baseline		Baseline		Änderung Woche 24 zu Baseline		Effektschätzer		
	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)			
	135	57,2 (17,61)	121	6,6 (18,84)	67	55,9 (17,85)	59	0,1 (13,58)	k. A.		n. b.

Studie PIONEER Darstellung der Ergebnisse <sup>1)</sup>	Avapritinib + BSC N = 141 <sup>2)</sup>				Placebo + BSC N = 71 <sup>2)</sup>				Wirkstoff vs. Komparator	Effekt
<b>Schwere Anaphylaxie-Episoden mit Einsatz von Epinephrin</b>	N <sup>3)</sup>	Personen mit Ereignis, n (%)			N <sup>3)</sup>	Personen mit Ereignis, n (%)			Effektschätzer	
	141	2 (1,4)			71	3 (4,2)			k. A.	n. b.
<b>Lebensqualität</b>										
<b>Gesundheitsbezogene Lebensqualität mittels SF-12</b>	Baseline		Änderung Woche 24 zu Baseline		Baseline		Änderung Woche 24 zu Baseline		Effektschätzer	
	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)		
Physical Component Summary <sup>13)</sup>	134	33,83 (10,79)	120	5,60 (8,43)	67	34,07 (10,70)	59	3,08 (7,01)	k. A.	n. b.
Mental Component Summary <sup>13)</sup>	134	40,60 (10,51)	120	3,92 (9,27)	67	40,99 (10,67)	59	1,62 (9,03)	k. A.	n. b.
<b>Krankheitsspezifische Lebensqualität mittels MC-QoL</b>	Baseline		Änderung Woche 24 zu Baseline		Baseline		Änderung Woche 24 zu Baseline		Effektschätzer	
	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)	N <sup>3)</sup>	MW (SD)		
Total Score <sup>14)</sup>	135	57,50 (16,02)	121	-19,20 (18,19)	68	57,47 (17,20)	60	-9,55 (14,93)	k. A.	n. b.
Symptome <sup>14)</sup>	135	65,14 (16,07)	121	-18,82 (19,18)	68	63,93 (17,55)	60	-10,32 (17,40)	k. A.	n. b.
Soziales Leben / Funktionieren <sup>14)</sup>	135	55,84 (21,02)	121	-20,09 (21,00)	68	55,33 (22,82)	60	-8,77 (18,82)	k. A.	n. b.
Emotionen <sup>14)</sup>	135	49,51 (23,39)	121	-18,73 (21,44)	68	48,98 (22,86)	60	-9,89 (17,83)	k. A.	n. b.
Haut <sup>14)</sup>	135	55,49 (22,33)	121	-18,60 (26,06)	68	61,89 (22,69)	60	-9,31 (18,80)	k. A.	n. b.
<b>Sicherheit</b>										
<b>Unerwünschte Ereignisse</b>	N <sup>15)</sup>	Personen mit Ereignis, n (%)			N <sup>15)</sup>	Personen mit Ereignis, n (%)			RR [95%-KI]; p-Wert	
Schwere UE	141	30 (21,3)			71	15 (21,1)			k. A.	n. b.
SUE	141	7 (5,0)			71	8 (11,3)			k. A.	n. b.
UE, das zum Abbruch der Studienmedikation führte	141	3 (2,1)			71	1 (1,4)			k. A.	n. b.

- 1) Ergebnisse beziehen sich auf die 24-wöchige RCT-Phase (Teil 2) der Studie PIONEER.
- 2) ITT-Population.
- 3) Die Anzahl entspricht denjenigen Personen, die für die Berechnung der jeweiligen Maßzahlen herangezogen wurden.
- 4) ANCOVA-Modell, das für den Randomisierungs-Stratifizierungsfaktor (Serumtryptase) und den Baseline-ISM-Status (mittelschwer vs. schwer) kontrolliert wurde, einseitiger p-Wert.
- 5) Skala von 0 bis 110. Je höher die Punktezahl, desto stärker ausgeprägt die Symptome.
- 6) Es liegen keine Angaben zu Hedges' g vor.
- 7) MWD mit 95%-KI und Zwei-Stichproben-t-Test.
- 8) Bei der Analyse der Domänenscores wurden die Randomisierungsstrata (Serumtryptase) und Baseline-ISM-Status (mittelschwer vs. schwer) nicht berücksichtigt. Dies war zwar präspezifiziert, ist methodisch jedoch nicht plausibel und nicht kongruent zum Gesamtscore.
- 9) Skala von 0 bis 30. Je höher die Punktezahl, desto stärker ausgeprägt die Symptome.
- 10) Skala von 0 bis 10. Je höher die Punktezahl, desto stärker ausgeprägt die Symptome.
- 11) Werte zwischen 0 (keine Symptome) und 4 (Symptome lassen sich nicht ignorieren).
- 12) Werte zwischen 0 (denkbar schlechtesten Gesundheitszustand) und 100 (denkbar besten Gesundheitszustand).
- 13) Werte zwischen 0 (denkbar schlechteste Lebensqualität) und 100 (denkbar beste Lebensqualität).
- 14) Werte zwischen 0 (geringe Beeinträchtigung der Lebensqualität) bis 100 (starke Beeinträchtigung der Lebensqualität).
- 15) Sicherheitspopulation.

Abkürzungen: ANCOVA: Kovarianzanalyse; BSC: Best Supportive Care; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ISM: Indolente systemische Mastozytose; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MC-QoL: Mastocytosis Quality of Life Questionnaire; LS: Least Squares; MW: Mittelwert; MWD: Mittelwertdifferenz; n. b.: nicht bewertbar; PGIS: Patient Global Impression of Severity; RCT: Randomisierte kontrollierte Studie; RR: Relatives Risiko; SD: Standardabweichung; SE: Standardfehler; SF-12: Short Form 12 Health Survey; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## Referenzen

1. **Blueprint Medicines.** A 3-part, randomized, double-blind, placebo-controlled phase 2 study to evaluate safety and efficacy of avapritinib (BLU-285), a selective KIT mutation-targeted tyrosine kinase inhibitor, in indolent and smoldering systemic mastocytosis with symptoms inadequately controlled with standard therapy; clinical protocol [unveröffentlicht]. 15.02.2018.
2. **Blueprint Medicines.** A 3-part, randomized, double-blind, placebo-controlled phase 2 study to evaluate safety and efficacy of avapritinib (BLU-285), a selective KIT mutation-targeted tyrosine kinase inhibitor, in indolent and smoldering systemic mastocytosis with symptoms inadequately controlled with standard therapy; clinical study report [unveröffentlicht]. 20.10.2022.
3. **Blueprint Medicines.** A 3-part, randomized, double-blind, placebo-controlled phase 2 study to evaluate safety and efficacy of avapritinib (BLU-285), a selective KIT mutation-targeted tyrosine kinase inhibitor, in indolent and smoldering systemic mastocytosis with symptoms inadequately controlled with standard therapy; statistical analysis plan [unveröffentlicht]. 15.02.2018.
4. **Blueprint Medicines.** Ayvakyt 25 mg, 50 mg, 100 mg 200 mg, 300 mg Filmtabletten [online]. 12.2023. Berlin. [Zugriff: 19.02.2024]. URL: <http://www.fachinfo.de>.
5. **Blueprint Medicines.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Avapritinib (AYVAKYT), Behandlung erwachsener Patienten mit indolenter systemischer Mastozytose (ISM) mit mittelschweren bis schweren Symptomen indiziert, bei denen mit einer symptomatischen Behandlung keine ausreichende Kontrolle erzielt werden kann; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 20.12.2023.
6. **European Medicines Agency (EMA).** Ayvakyt: European public assessment report EMEA/H/C/005208/II/0023 [online]. 09.11.2023. Amsterdam (NED): EMA. [Zugriff: 19.02.2004]. URL: [https://www.ema.europa.eu/en/documents/variation-report/ayvakyt-h-c-005208-ii-0023-epar-assessment-report-variation\\_en.pdf](https://www.ema.europa.eu/en/documents/variation-report/ayvakyt-h-c-005208-ii-0023-epar-assessment-report-variation_en.pdf).
7. **Gotlib J, Castells M, Elberink HO, Siebenhaar F, Hartmann K, Broesby-Olsen S, et al.** Avapritinib versus placebo in indolent systemic mastocytosis. *NEJM Evid* 2023;2(6):EVIDoa2200339.
8. **Guyatt GH, Norman GR, Juniper EF, Griffith LE.** A critical look at transition ratings. *J Clin Epidemiol* 2002;55(9):900-908.
9. **Jager U, Barcellini W, Broome CM, Gertz MA, Hill A, Hill QA, et al.** Diagnosis and treatment of autoimmune hemolytic anemia in adults: recommendations from the First International Consensus Meeting. *Blood Rev* 2020;41:100648.
10. **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.
11. **Norman GR, Stratford P, Regehr G.** Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50(8):869-879.

12. **Padilla B, Shields AL, Taylor F, Li X, McDonald J, Green T, et al.** Psychometric evaluation of the Indolent Systemic Mastocytosis Symptom Assessment Form (ISM-SAF) in a phase 2 clinical study. *Orphanet J Rare Dis* 2021;16(1):434.
13. **Shields AL, Taylor F, Lamoureux RE, Padilla B, Severson K, Green T, et al.** Psychometric evaluation of the Indolent Systemic Mastocytosis Symptom Assessment Form (ISM-SAF((c))) and determination of a threshold score for moderate symptoms. *Orphanet J Rare Dis* 2023;18(1):69.
14. **Siebenhaar F, von Tschirnhaus E, Hartmann K, Rabenhorst A, Staubach P, Peveling-Oberhag A, et al.** Development and validation of the mastocytosis quality of life questionnaire: MC-QoL. *Allergy* 2016;71(6):869-877.
15. **Taylor F, Akin C, Lamoureux RE, Padilla B, Green T, Boral AL, et al.** Development of symptom-focused outcome measures for advanced and indolent systemic mastocytosis: the AdvSM-SAF and ISM-SAF((c)). *Orphanet J Rare Dis* 2021;16(1):414.
16. **Ware J, Jr., Kosinski M, Keller SD.** A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34(3):220-233.

## Anhang

### WHO-Kriterien für die Diagnose und (Sub-)Klassifikation der systemischen Mastozytose

Diagnosis of SM: requires 1 major + 1 minor criterion <u>OR</u> 3 minor criteria	
Major Criteria	Minor Criteria
Multifocal dense infiltrates of MCs ( $\geq 15$ MCs in aggregates) in sections of BM and/or other extracutaneous organ(s)	In biopsy sections of BM or other extracutaneous organs, <ul style="list-style-type: none"> <li>&gt; 25% of all MCs in the infiltrate are spindle shaped or have atypical morphology or &gt; 25% of all MCs in BM aspirate are immature or atypical</li> <li>Detection of an activating point mutation at codon 816 of <i>KIT</i> in the BM, blood or another extracutaneous organ</li> <li>Mast cells in BM or blood or other extracutaneous organ express CD25 with or without CD2, in addition to normal MC markers<sup>a</sup></li> <li>Baseline serum total tryptase is persistently &gt; 20 ng/mL, unless there is an associated myeloid neoplasm, in which case this parameter is not valid.</li> </ul>
Classification of SM: Definition of B- and C-Findings	
B-Findings	C-Findings
<ul style="list-style-type: none"> <li>High MC burden (shown on BM biopsy): &gt; 30% infiltration of cellularity by MCs <i>and</i> serum total tryptase &gt; 200 ng/mL</li> <li>Signs of dysplasia or myeloproliferation in non-MC lineage(s), but criteria are not met for definitive diagnosis of an associated hematological neoplasms, with normal or only slightly abnormal blood counts.</li> <li>Hepatomegaly without impairment of liver function, palpable splenomegaly without hypersplenism and/or lymphadenopathy on palpation or imaging<sup>a</sup></li> </ul>	<ul style="list-style-type: none"> <li>Bone marrow dysfunction caused by neoplastic MC infiltration, manifested by <math>\geq 1</math> cytopenia:               <ul style="list-style-type: none"> <li>ANC &lt; <math>1.0 \times 10^9/L</math> or</li> <li>Hgb &lt; 10g/dL or</li> <li>Platelet count &lt; 100,000/<math>\mu L</math></li> </ul> </li> <li>Palpable hepatomegaly with impairment of liver function, ascites and/or portal hypertension<sup>b</sup></li> <li>Skeletal involvement, with large osteolytic lesions<sup>c</sup> with or without pathologic fractures (pathological fractures caused by osteoporosis do not qualify as a C-finding)</li> <li>Palpable splenomegaly with hypersplenism<sup>d</sup></li> <li>Malabsorption with weight loss due to gastrointestinal MC infiltrates<sup>e</sup></li> </ul>

Abbreviations: ANC = absolute neutrophil count; BM = bone marrow; CD2 = cluster of differentiation 2; CD25 = cluster of differentiation 25; CT = computed tomography; Hgb = hemoglobin; *KIT* = V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; MC = mast cell; MDS = myelodysplastic syndrome; MPN = myeloproliferative neoplasm; SM = systemic mastocytosis; WHO = World Health Organization.

<sup>a</sup> Organomegaly may be documented by palpation or imaging. Lymphadenopathy is defined as > 2 cm by CT or ultrasound.

<sup>b</sup> Signs of impaired liver function include elevated transaminases and/or bilirubin levels and/or hyperalbuminemia (with or without ascites or portal hypertension).

<sup>c</sup> Large osteolytic lesions are defined as  $\geq 2$  cm.

<sup>d</sup> Hypersplenism is defined as splenomegaly in association with a cytopenia (eg, platelets < 100,000/ $\mu L$ ).

<sup>e</sup> Malabsorption with weight loss is defined as hypoalbuminemia or weight loss > 10% with a 6-month period.

Abbildung 1: WHO-Kriterien für die Diagnose und Klassifikation der systemischen Mastozytose [1]



WHO Diagnosis	Criteria
ISM	No C-findings; < 2 B-findings
SSM	No C-findings; ≥ 2 B-findings
SM-AHN	Diagnostic criteria for MDS or MPN
ASM	≥ 1 C-finding
MCL	≥ 20% MCs in BM smear

Abbreviations: ASM = aggressive systemic mastocytosis; BM = bone marrow; ISM = indolent systemic mastocytosis; MC = mast cells; MCL = mast cell leukemia; MDS = myelodysplastic syndrome; MPN = myeloproliferative neoplasm; SM-AHN = systemic mastocytosis with an associated hematologic neoplasm of non-mast-cell lineage; SSM = smoldering systemic mastocytosis; WHO = World Health Organization.

Abbildung 2: WHO-Kriterien für die Sub-Klassifikation der systemischen Mastozytose [1]

## Ergänzende Ergebnisse

### ISM-SAF

Tabelle 24: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite des ISM-SAF zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022)

Studie PIONEER ISM-SAF Responderanalyse zu Woche 24	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Gesamt-Symptomscore (Verbesserung um mind. 15 % der Skalenspannbreite (≥ 16,5 Punkte))</b>		
Responder <sup>1)</sup> , n (%)	54 (43,9)	19 (28,4)
RR [95%-KI]; p-Wert <sup>2)</sup>	1,50 [0,98; 2,31]; 0,065	
<b>Gastrointestinaler Symptomscore (≥ 4,5 Punkte)<sup>3)</sup></b>		
Responder <sup>1)</sup> , n (%)	45 (36,6)	25 (37,3)
RR [95%-KI]; p-Wert <sup>2)</sup>	0,98 [0,67; 1,42]; > 0,999	
<b>Haut-Symptomscore (≥ 4,5 Punkte)<sup>3)</sup></b>		
Responder <sup>1)</sup> , n (%)	70 (56,9)	19 (28,4)
RR [95%-KI]; p-Wert <sup>2)</sup>	1,97 [1,30; 2,98]; 0,001	
<b>Neurokognitives Symptomcluster (≥ 4,5 Punkte)<sup>3)</sup></b>		
Responder <sup>1)</sup> , n (%)	47 (38,2)	20 (29,9)
RR [95%-KI]; p-Wert <sup>2)</sup>	1,25 [0,82; 1,91]; 0,296	

<sup>1)</sup> Der Baseline-Score ist definiert als der 14-Tage-Durchschnitt der 14 Tage vor der ersten Einnahme der Studienmedikation. Der Wert zu Zyklus 7 Tag 1 ist der 14-Tage-Durchschnitt vor Zyklus 7 Tag 1. Wenn bei einer Person mehr als 7 Tage (Werte) fehlen, wird der Score für die Person als fehlend betrachtet. Personen mit fehlendem Score zu Baseline oder zu Zyklus 7 Tag 1 werden im Nenner gezählt.

Personen mit hochdosierter Steroideinnahme (> 20 mg/d Prednison oder Äquivalent) innerhalb von 7 Tagen vor Zyklus 7 Tag 1 oder an mehr als 14 aufeinanderfolgenden Tagen zu irgendeinem Zeitpunkt zwischen Baseline und Zyklus 7 Tag 1 werden im Nenner, aber nicht im Zähler gezählt.

<sup>2)</sup> Zweiseitiger p-Wert des RR basiert auf der asymptotischen normalen Annäherung der logarithmierten Mantel-Haenszel-Schätzung, stratifiziert nach dem Randomisierungs-Stratifizierungsfaktor (Serumtryptase < 20 ng/ml vs. ≥ 20 ng/ml) und dem Baseline-ISM-Status (mittelschwer vs. schwer).

<sup>3)</sup> Entspricht 15 % der Skalenspannbreite des ISM-SAF.

Abkürzungen: BSC: Best Supportive Care; ISM: Indolente systemische Mastozytose; ISM-SAF: Indolent Systemic Mastocytosis Symptom Assessment Form; KI: Konfidenzintervall; PP: Per Protokoll; RR: Relatives Risiko.

PGIS

*Tabelle 25: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite des PGIS zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022)*

Studie PIONEER PGIS Responderanalyse zu Woche 24	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Verbesserung um mind. 15 % der Skalenspannbreite (<math>\geq 1</math> Punkt)</b>		
Responder, n (%)	65 (52,8)	23 (34,3)
RR [95%-KI]; p-Wert <sup>1)</sup>	1,61 [1,12; 2,33]; 0,011	

<sup>1)</sup> Zweiseitiger p-Wert des RR basiert auf der asymptotischen normalen Annäherung der logarithmierten Mantel-Haenszel-Schätzung, stratifiziert nach dem Randomisierungs-Stratifizierungsfaktor (Serumtryptase < 20 ng/ml vs.  $\geq 20$  ng/ml) und dem Baseline-ISM-Status (mittelschwer vs. schwer).

Abkürzungen: BSC: Best Supportive Care; ISM: Indolente systemische Mastozytose; KI: Konfidenzintervall; PGIS: Patient Global Impression of Severity; PP: Per Protokoll; RR: Relatives Risiko.

EQ-5D-VAS

*Tabelle 26: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite der EQ-5D-VAS zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022)*

Studie PIONEER EQ-5D-VAS Responderanalyse zu Woche 24	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Verbesserung um mind. 15 % der Skalenspannbreite</b>		
Responder, n (%)	37 (30,1)	6 (9,0)
RR [95%-KI]; p-Wert <sup>1)</sup>	3,42 [1,53; 7,63]; 0,003	

<sup>1)</sup> Zweiseitiger p-Wert des RR basiert auf der asymptotischen normalen Annäherung der logarithmierten Mantel-Haenszel-Schätzung, stratifiziert nach dem Randomisierungs-Stratifizierungsfaktor (Serumtryptase < 20 ng/ml vs.  $\geq 20$  ng/ml) und dem Baseline-ISM-Status (mittelschwer vs. schwer).

Abkürzungen: BSC: Best Supportive Care; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ISM: Indolente systemische Mastozytose; KI: Konfidenzintervall; PP: Per Protokoll; RR: Relatives Risiko.

PGIC – Grad der Veränderung

*Tabelle 27: Mittlere Veränderung des Gesundheitszustands mittels PGIC; Studie PIONEER, ITT-Population (Datenschnitt: 23.06.2022)*

Studie PIONEER PGIC – Grad der Veränderung	Avapritinib + BSC N = 141	Placebo + BSC N = 71
<b>Mittlere Veränderung</b>		
Veränderung zu Woche 24 n (%)	127	62
MW (SD)	3,6 (1,81)	4,9 (1,54)
MWD [95%-KI]; p-Wert	k. A.	

<sup>1)</sup> Werte zwischen 0 (große Verbesserung) und 10 (große Verschlechterung).

Abkürzungen: BSC: Best Supportive Care; ITT: Intention-to-Treat; k. A.: keine Angabe; KI: Konfidenzintervall; MW: Mittelwert; MWD: Mittelwertdifferenz; PGIC: Patient Global Impression of Change; SD: Standardabweichung.

**Tabelle 28: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannbreite des PGIC zu Woche 24; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022)**

Studie PIONEER PGIC – Grad der Veränderung Responderanalyse zu Woche 24	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Verbesserung um mind. 15 % der Skalenspannbreite (≥ 1,5 Punkte)</b>		
Responder, n (%)	9 (7,3)	1 (1,5)
RR [95%-KI]; p-Wert <sup>1)</sup>	4,69 [0,50; 43,81]; 0,175	

<sup>1)</sup> Zweiseitiger p-Wert des RR basiert auf der asymptotischen normalen Annäherung der logarithmierten Mantel-Haenszel-Schätzung, stratifiziert nach dem Randomisierungs-Stratifizierungsfaktor (Serumtryptase < 20 ng/ml vs. ≥ 20 ng/ml) und dem Baseline-ISM-Status (mittelschwer vs. schwer).

Abkürzungen: BSC: Best Supportive Care; ISM: Indolente systemische Mastozytose; KI: Konfidenzintervall; PGIC: Patient Global Impression of Change; PP: Per Protokoll; RR: Relatives Risiko.

### Gesundheitsbezogene Lebensqualität mittels SF-12

**Tabelle 29: Gesundheitsbezogene Lebensqualität mittels SF-12; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022)**

Studie PIONEER SF-12	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Physical Component Summary<sup>1)</sup></b>		
<i>Baseline</i> n (%) MW (SD)	118 (95,9) 33,16 (10,59)	64 (95,5) 34,28 (10,90)
<i>Veränderung zu Woche 24</i> n (%) LS Mean (SE) <sup>2)</sup>	106(86,2) 4,35 (0,89)	56 (83,6) 0,81 (1,16)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	3,55 [1,11; 5,98]; 0,005	
Hedges' g [95%-KI]	0,39 [0,07; 0,73]	
<b>Mental Component Summary<sup>1)</sup></b>		
<i>Baseline</i> n (%) MW (SD)	118 (95,9) 39,93 (10,05)	64 (95,5) 40,61 (10,51)
<i>Veränderung zu Woche 24</i> n (%) LS Mean (SE) <sup>2)</sup>	106 (86,2) 4,03 (1,10)	56 (83,6) 1,51 (1,44)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	2,53 [-0,51; 5,57]; 0,103	

<sup>1)</sup> Werte zwischen 0 (denkbar schlechteste Lebensqualität) und 100 (denkbar beste Lebensqualität).

<sup>2)</sup> ANCOVA-Modell, das für den Randomisierungs-Stratifizierungsfaktor (Serumtryptase) und den Baseline-ISM-Status (mittelschwer vs. schwer) kontrolliert wurde, zweiseitiger p-Wert.

Abkürzungen: ANCOVA: Kovarianzanalyse; BSC: Best Supportive Care; ISM: Indolente systemische Mastozytose; KI: Konfidenzintervall; LS: Least Squares; MW: Mittelwert; PP: Per Protokoll; SD: Standardabweichung; SE: Standardfehler; SF-12: Short Form 12 Health Survey;

Krankheitsspezifische Lebensqualität mittels MC-QoL

Tabelle 30: Lebensqualität mittels MC-QoL; Studie PIONEER, PP-Population (Datenschnitt: 23.06.2022)

Studie PIONEER MC-QoL	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Gesamtscore<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	119 (96,8)	65 (97,0)
MW (SD)	59,26 (15,46)	58,26 (16,98)
<i>Veränderung zu Woche 24</i>		
n (%)	107 (87,0)	57 (85,1)
LS Mean (SE) <sup>2)</sup>	-17,11 (1,97)	-6,31 (2,55)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	-10,80 [-16,26; -5,34]; < 0,001	
Hedges' g [95%-KI]	-0,54 [-0,88; -0,22]	
<b>Symptome<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	119 (96,8)	65 (97,0)
MW (SD)	67,09 (14,75)	64,23 (17,83)
<i>Veränderung zu Woche 24</i>		
n (%)	107 (87,0)	57 (85,1)
LS Mean (SE) <sup>2)</sup>	-16,92 (2,08)	-6,57 (2,70)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	-10,35 [-16,14; -4,56]; < 0,001	
Hedges' g [95%-KI]	-0,49 [-0,82; -0,17]	
<b>Soziales Leben / Funktionieren<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	119 (96,8)	65 (97,0)
MW (SD)	57,33 (21,17)	56,13 (22,79)
<i>Veränderung zu Woche 24</i>		
n (%)	107 (87,0)	57 (85,1)
LS Mean (SE) <sup>2)</sup>	-17,40 (2,39)	-5,31 (3,10)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	-12,09 [-18,73; -5,44]; < 0,001	
Hedges' g [95%-KI]	-0,50 [-0,83; -0,17]	
<b>Emotionen<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	119 (96,8)	65 (97,0)
MW (SD)	51,44 (22,92)	50,35 (22,24)
<i>Veränderung zu Woche 24</i>		
n (%)	107 (87,0)	57 (85,1)
LS Mean (SE) <sup>2)</sup>	-17,25 (2,36)	-7,58 (3,06)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	-9,67 [-16,24; -3,11]; 0,004	
Hedges' g [95%-KI]	-0,40 [-0,73; -0,08]	

Studie PIONEER MC-QoL	Avapritinib + BSC N = 123	Placebo + BSC N = 67
<b>Haut<sup>1)</sup></b>		
<i>Baseline</i>		
n (%)	119 (96,8)	65 (97,0)
MW (SD)	57,21 (22,02)	62,95 (22,39)
<i>Veränderung zu Woche 24</i>		
n (%)	107 (87,0)	57 (85,1)
LS Mean (SE) <sup>2)</sup>	-16,47 (2,70)	-6,39 (3,50)
LS Mean difference [95%-KI]; p-Wert <sup>2)</sup>	-10,08 [-17,58; -2,58]; 0,009	
Hedges' g [95%-KI]	-0,37 [-0,70; -0,04]	

<sup>1)</sup> Werte zwischen 0 (geringe Beeinträchtigung der Lebensqualität) bis 100 (starke Beeinträchtigung der Lebensqualität).

<sup>2)</sup> ANCOVA-Modell, das für den Randomisierungs-Stratifizierungsfaktor (Serumtrypsinase) und den Baseline-ISM-Status (mittelschwer vs. schwer) kontrolliert wurde, 2-seitiger p-Wert.

Abkürzungen: ANCOVA: Kovarianzanalyse; BSC: Best Supportive Care; ISM: Indolente systemische Mastozytose; KI: Konfidenzintervall; MC-QoL: Mastocytosis Quality of Life Questionnaire; LS: Least Squares; MW: Mittelwert; PP: Per Protokoll; SD: Standardabweichung; SE: Standardfehler.

### Sicherheit – Unerwünschte Ereignisse

*Tabelle 31: Gesamtraten der UE bis Woche 24; Studie PIONEER, PP-Population  
(Datenschnitt: 23.06.2022)*

Studie PIONEER Gesamtraten der UE bis Woche 24 Personen mit mindestens einem ...	Avapritinib + BSC N = 123 n (%)	Placebo + BSC N = 67 n (%)	RR <sup>1)</sup> [95%-KI]; p-Wert
UE (ergänzend dargestellt)	114 (92,7)	62 (92,5)	-
UE CTCAE-Grad ≥ 3	26 (21,1)	14 (20,9)	1,012 [0,568; 1,802]; 0,969
SUE	6 (4,9)	8 (11,9)	0,409 [0,148; 1,128]; 0,084
UE, das zum Abbruch der Studienmedikation führte	1 (0,8)	1 (1,5)	0,545 [0,035; 8,570]; 0,666

<sup>1)</sup> Zwei-Stichproben-t-Test.

Abkürzungen: BSC: Best Supportive Care; CTCAE: Common Terminology Criteria for Adverse Events; KI: Konfidenzintervall; PP: Per Protokoll; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis; - : nicht berechenbar.