

**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: Vimseltinib**

Datum der Veröffentlichung: 2. Februar 2026

Version 1.0

## 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 .....	9
2.3 Endpunkte .....	13
2.3.1 Mortalität.....	14
2.3.2 Morbidität.....	14
2.3.3 Lebensqualität .....	24
2.3.4 Sicherheit .....	24
2.3.5 Erhebungszeitpunkte .....	26
2.4 Statistische Methoden.....	27
2.5 Verzerrungspotential auf Studien- und Endpunktebene .....	29
3 Ergebnisse der eingeschlossenen Studie .....	31
3.1 Studiencharakteristika und Studienmedikation .....	31
3.2 Mortalität .....	34
3.3 Morbidität .....	35
3.4 Lebensqualität.....	39
3.5 Sicherheit .....	39
4 Diskussion der Methodik und Ergebnisse.....	43
4.1 Design und Methodik der Studie.....	43
4.2 Zulassungsstatus und Zulassungspopulation.....	44
4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit .....	45
4.4 Mortalität .....	46
4.5 Morbidität .....	46
4.6 Lebensqualität.....	47
4.7 Sicherheit .....	47
5 Anforderung an eine qualitätsgesicherte Anwendung .....	49
6 Zusammenfassung der Nutzenbewertung .....	50
Referenzen .....	53
Anhang .....	55

## Tabellenverzeichnis

Tabelle 1:	Übersicht über die Studienbasis .....	8
Tabelle 2:	Charakterisierung der Studie MOTION.....	9
Tabelle 3:	Für die Nutzenbewertung relevante Protokolländerungen der Studie MOTION.....	12
Tabelle 4:	Charakterisierung der Intervention in der Studie MOTION .....	12
Tabelle 5:	Zusammenfassung der Endpunktbewertung der Studie MOTION.....	13
Tabelle 6:	Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie MOTION .....	26
Tabelle 7:	Verzerrungspotential der Studie MOTION.....	29
Tabelle 8:	Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie MOTION .....	30
Tabelle 9:	Allgemeine Angaben; Studie MOTION .....	31
Tabelle 10:	Charakterisierung der Studienpopulation; Studie MOTION .....	32
Tabelle 11:	Begleitmedikation bei mindestens 15 % in einem Studienarm; Studie MOTION, Teil 1 .....	34
Tabelle 12:	PROMIS PF, MMRM-Analysen anhand beobachteter Werte zu Woche 25; Studie MOTION, ITT-Population .....	35
Tabelle 13:	Anteil der Personen mit einer Verbesserung um mindestens 2 Punkte der NRS-Steifigkeit zu Woche 25; Studie MOTION, ITT-Population .....	36
Tabelle 14:	Schmerzen mittels BPI-SF, MMRM-Analysen anhand beobachteter Werte zu Woche 25; Studie MOTION, ITT-Population .....	36
Tabelle 15:	Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalen- spannweite der EQ-5D-VAS zu Woche 25; Studie MOTION, ITT-Population .....	38
Tabelle 16:	Anteil der Personen mit einer Verbesserung im PGIS zu Woche 25 (ergänzend dargestellt); Studie MOTION, ITT-Population .....	38
Tabelle 17:	Anteil der Personen mit einer Verbesserung im PGIC-Krankheitssymptomatik zu Woche 25; Studie MOTION, ITT-Population .....	39
Tabelle 18:	Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase, Studie MOTION, Sicherheitspopulation .....	40
Tabelle 19:	UE mit Inzidenz $\geq 10$ % während der verblindeten Behandlungsphase; Studie MOTION, Sicherheitspopulation .....	40
Tabelle 20:	Schwere UE (CTCAE-Grad $\geq 3$ ), die bei $\geq 5$ % der Personen in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term, während der verblindeten Behandlungsphase; Studie MOTION, Sicherheitspopulation .....	42
Tabelle 21:	Zusammenfassende Darstellung der Ergebnisse der Studie MOTION.....	50
Tabelle 22:	Bewegungsumfang, MMRM-Analysen anhand beobachteter Werte zu Woche 25; Studie MOTION, ITT-Population .....	55
Tabelle 23:	Anteil der Personen mit einer Verbesserung im PGIC zu Woche 25; Studie MOTION, ITT-Population .....	56

Tabelle 24: Anteil der Personen mit Tumoransprechen zu Woche 25; Studie MOTION, ITT-Population .....	57
Tabelle 25: Faktoren des Gesamtansprechens gemäß RECIST (Version 1.1) .....	57

## Abkürzungsverzeichnis

AMA	American Medical Association
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
API	Application Programming Interface
BPI-SF	Brief Pain Inventory – Short Form
CDC	Centers for Disease Control and Prevention
CR	Komplettes Ansprechen (Complete Response)
CSF1(R)	Colony Stimulating Factor 1 (Receptor)
CTCAE	Common Terminology Criteria for Adverse Events
EPAR	European Public Assessment Report
EQ-5D-VAS	Visuelle Analogskala des European Quality of Life 5-Dimension
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IRR	Unabhängige radiologische Überprüfung (Independent Radiological Review)
ITT	Intention to Treat
KI	Konfidenzintervall
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measures
mRECIST	modified Response Evaluation Criteria In Solid Tumors
MRT	Magnetresonanztomographie
NRS	Numerische Bewertungsskala (Numeric Rating Scale)
ORR	Objektive Ansprechrate (Objective Response Rate)
PF	Körperliche Funktionsfähigkeit (Physical Function)
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PR	Partielles Ansprechen (Partial Response)
PROMIS PF	Patient-Reported Outcomes Measurement Information System Physical Function
PT	Preferred Term
pU	pharmazeutischer Unternehmer
RECIST	Response Evaluation Criteria In Solid Tumors
ROM	Bewegungsumfang (Range of Motion)
RR	Relatives Risiko
SAP	Statistischer Analyseplan
SGB	Sozialgesetzbuch
SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis
TGCT	Tenosynovialer Riesenzelltumor (Tenosynovial Giant Cell Tumour)
TVS	Tumorvolumen-Score (Tumor Volume Score)
UE	Unerwünschtes Ereignis
VerfO	Verfahrensordnung des G-BA
WHO	World Health Organization

## Hintergrund

Vimseltinib 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 Vimseltinib 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 Vimseltinib in seiner Sitzung am 27. Januar 2026 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 30. Oktober 2025 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pU in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 2. Februar 2026 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

Vimseltinib (Romvimza®) 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 [5]:

Erwachsene Patientinnen und Patienten mit symptomatischen tenosynovialen Riesenzell-tumoren (Tenosynovial Giant Cell Tumours, TGCT), die mit einer klinisch relevanten Verschlechterung der körperlichen Funktionsfähigkeit assoziiert sind und bei denen chirurgische Optionen ausgeschöpft sind oder zu einer inakzeptablen Morbidität oder Behinderung führen würden.

Gemäß Fachinformation beträgt die zugelassene Dosierung 30 mg zweimal wöchentlich im Abstand von mindestens 72 Stunden, wobei die Therapie so lange fortgesetzt werden kann, wie die Patientin / der Patient davon einen Nutzen hat und keine inakzeptable Toxizität auftritt.

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

*Tabelle 1: Übersicht über die Studienbasis*

Studienname (Nummer)	Studie mit Dossier vom pU eingereicht	Studie vom pU als relevant für die Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzen- bewertung	Ausschlussgrund
<b>Studie zum Wirkstoff</b>				
MOTION (DCC-3014-03-001) <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 denen die Einschätzungen des pU zum Zusatznutzen basiert.

#### **Zur Nutzenbewertung für Vimseltinib herangezogene Unterlagen**

- Dossier des pU zu Vimseltinib [6]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [8]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der Studie MOTION (DCC-3014-03-001) [4]
- Fachinformation zu Vimseltinib [5]
- Publikationen zum Wirkstoff: Gelderblom et al. (2024) [10]



## 2.2 Beschreibung der eingeschlossenen Studie

Die Nutzenbewertung von Vimseltinib im vorliegenden Anwendungsgebiet basiert auf der Zulassungsstudie MOTION (DCC-3014-03-001). Die Studie und die Intervention werden in Tabelle 2 und Tabelle 4 charakterisiert. Für die vorliegende Nutzenbewertung werden ausschließlich die Ergebnisse des randomisierten, vergleichenden Studienteils 1 als relevant erachtet.

*Tabelle 2: Charakterisierung der Studie MOTION*

Charakteristikum	Beschreibung
<b>Design und Studienablauf</b>	<p>MOTION ist eine randomisierte, multizentrische, Placebo-kontrollierte Studie zur Untersuchung der Wirksamkeit und Sicherheit von Vimseltinib bei Personen mit TGCT bestehend aus einem doppelblinden Teil 1 und einem offenen, einarmigen Teil 2. Randomisiert wurde im Verhältnis 2:1 auf die Behandlungsarme Vimseltinib oder Placebo, stratifiziert nach Tumorlokalisation (untere Gliedmaße vs. alle andere) und Region (USA vs. Rest der Welt).</p> <p>Nach Woche 24 können die Personen des Placebo-Arms in den Vimseltinib-Arm wechseln. Die Studie lässt sich in folgende Abschnitte unterteilen:</p> <ul style="list-style-type: none"> <li>• Screening (maximal 42 Tage vor Behandlungsbeginn)</li> <li>• Behandlungsperiode Teil 1 (bis Woche 25)</li> <li>• Behandlungsperiode Teil 2 (bis Woche 49)</li> <li>• Erweiterungsperiode nach Woche 49</li> <li>• Sicherheits-Nachbeobachtung (30 Tage nach letzter Studienmedikation)</li> <li>• Follow-up für Krankheitsstatus für bis zu 2 Jahre (Personen ohne radiologische Krankheitsprogression oder Rückzug der Einverständniserklärung: Visite 12 Wochen nach letzter Dosis; danach telefonisch alle 16 Wochen inkl. Review der Krankenakte)</li> </ul> <p>Die Studie ist laufend und das Studienende ist für den Juli 2026 geplant.</p> <p>Initial in den Placebo-Arm randomisierte Personen können in Teil 2 der Studie Vimseltinib erhalten und initial in den Vimseltinib-Arm randomisierte Personen werden auch in Teil 2 der Studie weiterhin mit Vimseltinib behandelt. Die Behandlung erfolgt bis Krankheitsprogression gemäß verblindetem IRR oder Studienende, je nachdem welches Ereignis zuerst eintritt. Kommt es basierend auf der Bewertung durch die IRR zu einer Krankheitsprogression vor Beendigung des Teil 1 der Studie, wird die Testperson entblindet. Personen aus dem Placebo-Arm mit Krankheitsprogression können frühzeitig in Teil 2 übergehen und unverblindet Vimseltinib erhalten.</p> <p>Neben den Ergebnissen des abgeschlossenen, randomisierten Studienteils 1 legt der pU auch Daten des laufenden, nichtvergleichenden Studienteils 2 mit einem Datenschnitt vom 22.02.2025 vor.</p> <p>Für die vorliegende Nutzenbewertung werden ausschließlich die Ergebnisse des randomisierten Studienteil 1 als relevant erachtet.</p>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Alter <math>\geq 18</math> Jahre.</li> <li>• Histologisch bestätigte Diagnose von TGCT; in einem einzigen Gelenk, in dem eine Bewertung des Bewegungsumfangs vorgenommen werden kann.</li> <li>• Erkrankung, bei der gemäß der Beurteilung im Rahmen einer chirurgischen Beratung oder einem multidisziplinären Tumorboard eine chirurgische Resektion möglicherweise zu einer Verschlechterung der Funktionseinschränkung oder einer schweren Morbidität führen würde.</li> <li>• Symptomatische Erkrankung mit mind. mäßigen Schmerzen oder mind. mäßiger Steifigkeit (definiert als ein Wert von 4 oder mehr, wobei 10 den schlechtesten</li> </ul>

Charakteristikum	Beschreibung
	<p>Zustand beschreibt) innerhalb des Screening-Zeitraums (Fragebögen sollten an 14 aufeinanderfolgenden Tagen ausgefüllt werden).</p> <ul style="list-style-type: none"> <li>• Stabiles Analgetika-Regime mind. 2 Wochen vor der ersten Dosis.</li> <li>• Messbare Erkrankung gemäß RECIST mit mind. einer Läsion von mind. 2 cm Größe (MRT-Aufnahmen durch zentralen Radiologen beurteilt).</li> <li>• Angemessene Organfunktion und Knochenmarkreserve gemäß den folgenden Laboruntersuchungen: <ul style="list-style-type: none"> <li>◦ Knochenmarkfunktion: Absolute Neutrophilenzahl <math>\geq 1.500/\mu\text{l}</math>; Hämoglobin <math>\geq 10 \text{ g/dl}</math>; Thrombozytenzahl <math>\geq \text{LLN}</math>.</li> <li>◦ Leberfunktion: Gesamtserumbilirubin <math>\leq \text{ULN}</math>; AST/ALT <math>\leq \text{ULN}</math>.</li> <li>◦ Nierenfunktion: Kreatinin-Clearance <math>\geq 50 \text{ ml/min}</math>.</li> <li>◦ Elektrolyte <math>\geq \text{LLN}</math> für: Kalium, Magnesium und Kalzium.</li> </ul> </li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Vorherige Anwendung einer systemischen Therapie, die auf CSF1 oder CSF1R abzielt; eine vorherige Therapie mit Imatinib oder Nilotinib ist erlaubt.</li> <li>• Bekannter metastasierter TGCT oder anderer aktiver Krebs, der eine gleichzeitige Behandlung erfordert.</li> <li>• Verlängerung des QTcF-Intervalls auf der Grundlage eines wiederholten Nachweises von QTcF <math>&gt; 450 \text{ ms}</math> bei Männern oder <math>&gt; 470 \text{ ms}</math> bei Frauen oder einer Vorgeschichte mit langem QT-Syndrom.</li> <li>• Gleichzeitige Behandlung mit nicht erlaubten Medikamenten.</li> <li>• Klinisch bedeutsame Begleiterkrankungen.</li> <li>• Aktive Leber- oder Gallenerkrankung einschl. NASH oder Zirrhose.</li> </ul>
<b>Intervention und Zahl der Patientinnen und Patienten</b>	<p>Gescreent gesamt: N = 146  Randomisiert/Eingeschlossen: N = 123  Intervention: N = 83  Kontrolle: N = 40</p>
<b>Ort und Zeitraum der Durchführung; Datenschnitte</b>	<p><b>Studienzentren</b>  30 Studienzentren in 13 Ländern: USA, Italien, Spanien, Frankreich, Deutschland, Vereinigtes Königreich, Australien, Kanada, Hongkong, Niederlande, Norwegen, Polen, Schweiz.</p> <p><b>Studienzeitraum</b></p> <ul style="list-style-type: none"> <li>• Erste Person erste Visite: 15.10.2021</li> <li>• Abschluss der doppelblinden Behandlungsphase: 22.08.2023</li> <li>• Geplantes Studienende: Juli 2026</li> </ul> <p><b>Datenschnitt</b></p> <ul style="list-style-type: none"> <li>• Abschluss der doppelblinden Behandlungsphase: 22.08.2023</li> <li>• Laufende einarmigen Studienphase: 22.02.2025</li> </ul>

Charakteristikum	Beschreibung
<b>Endpunkte gemäß Studienprotokoll</b>	<p><b>Primärer Endpunkt</b> Objektive Ansprechrate (ORR, definiert als CR und PR) nach RECIST</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• ORR gemäß TVS und mRECIST</li> <li>• Veränderung des aktiven Bewegungsumfangs des betroffenen Gelenks</li> <li>• PROMIS PF</li> <li>• NRS für die Steifigkeit</li> <li>• VAS des EQ-5D-5L</li> <li>• Stärkste Schmerzen gemäß Item des BPI-SF</li> <li>• ORR gemäß mRECIST</li> <li>• Dauer des Ansprechens</li> <li>• Sicherheit</li> </ul> <p><b>Explorative Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Gewebeschaden-Skalenwert (Tissue Damage Score, TDS)</li> <li>• Durchschnittliche Schmerzen gemäß Item des BPI-SF</li> <li>• PGIC, CGIC<sup>1)</sup></li> <li>• PGIS und CGIS<sup>1)</sup></li> <li>• GP5-Frage des FACT-G</li> </ul>

<sup>1)</sup> Die Fragen zum CGIC/CGIS entsprechen denen des PGIS/PGIC und der pU berücksichtigt die Ergebnisse nicht in Modul 4 des Dossiers. Im vorliegenden Anwendungsgebiet werden die von Patientinnen und Patienten selbst eingeschätzten Endpunkte bevorzugt. Analog zum pU werden die Ergebnisse des CGIC/CGIS nicht zur Ableitung des Zusatznutzens berücksichtigt.

Abkürzungen: AST: Aspartat-Aminotransferase; ALT: Alanin-Aminotransferase; BPI-SF: Brief Pain Inventory – Short Form; CGIC: Clinical Global Impression of Change; CGIS: Clinical Global Impression of Severity; CR: Komplettes Ansprechen; CSF1(R): Colony Stimulating Factor 1 (Receptor); EQ-5D-5L: European Quality of Life 5-Dimension 5-Level; FACT-G: Functional Assessment of Cancer Therapy – General; IRR: Unabhängige radiologische Überprüfung; LLN: Lower Limit of Normal; mRECIST: modified Response Evaluation Criteria In Solid Tumors; MRT: Magnetresonanztomographie; NASH: Nichtalkoholische Steatohepatitis; NRS: Numerische Bewertungsskala; ORR: Objektive Ansprechrate; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PR: Partielles Ansprechen; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; pU: pharmazeutischer Unternehmer; RECIST: Response Evaluation Criteria In Solid Tumors; TGCT: Tenosynovialer Riesenzelltumor; TVS: Tumolvolumen-Score; ULN: Upper Limit of Normal; VAS: Visuelle Analogskala.

## Protokolländerungen

Es wurden 3 Änderungen des Originalprotokolls vom 15.12.2020 vorgenommen, wovon 2 nach Einschluss der ersten Person in die Studie am 15.10.2021 implementiert wurden. Bei der Version 2 waren bereits 81 Personen in die Studie eingeschlossen, zum Zeitpunkt von Version 3 bereits 116 Personen.

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

Amendment	Wesentliche Änderungen
Version 2 vom 07.03.2022 (Anzahl bis dahin eingeschlossene Personen: n = 81)	<ul style="list-style-type: none"> <li>• Festlegung, dass die Teilnehmenden TGCT in einem einzelnen Gelenk sowie in Gelenken haben müssen, bei denen der Bewegungsumfang im Gelenk beurteilt werden kann. Somit soll die Teilnahme von Personen gesichert werden, für die eine Bewertung des Bewegungsumfangs vorgenommen werden kann, da die Bewertung einen der sekundären Endpunkte darstellt.</li> <li>• Klarstellung, dass nur eine Zielläsion mind. 2 cm groß sein muss, um als geeignet zu gelten.</li> <li>• Anpassungen bezüglich unzulässiger Medikamente und Therapien.</li> <li>• Aufnahme von Empfehlungen zum Umgang mit Bilirubin-Erhöhen.</li> <li>• Anpassung der Kriterien für Dosisunterbrechungen bei hepatobiliären Laborwertsteigerungen, ALT- und/oder AST-Erhöhen sowie Gesamtbilirubin und behandlungsbedingten, verlängerten QT-Intervallen.</li> <li>• Angabe von Empfehlungen für den Umgang mit bestimmten behandlungsbedingten dermatologischen Nebenwirkungen.</li> <li>• Erklärung, dass Krankheitsprogression nicht als UE gilt, um die Leitlinie zur Meldung schwerwiegender Nebenwirkungen zu präzisieren.</li> </ul>
Version 3 vom 31.05.2023 (Anzahl bis dahin eingeschlossene Personen: n = 116)	<ul style="list-style-type: none"> <li>• Anpassung der Kriterien für Dosisunterbrechungen.</li> <li>• Anpassung des Betrachtungszeitraums für das Item „Stärkster Schmerz“ des BPI-SF zu Baseline von 7 auf 14 Tage.</li> </ul>

Abkürzungen: AST: Aspartat-Aminotransferase; ALT: Alanin-Aminotransferase; BPI-SF: Brief Pain Inventory – Short Form; TGCT: Tenosynovialer Riesenzelltumor; UE: Unerwünschtes Ereignis.

## Charakterisierung der Intervention

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

Intervention	Kontrolle
<b>Vimseltinib</b> <ul style="list-style-type: none"> <li>• Dosierung: 30 mg zweimal wöchentlich oral an Tag 1 und Tag 5, in Zyklen bestehend aus 28 Tagen.</li> <li>• Einnahme zur gleichen Tageszeit auf nüchternen Magen, mind. 1 Stunde vor und frühestens 2 Stunden nach der Nahrungsaufnahme.</li> <li>• Dosisreduktionen in 2 Schritten auf 20 mg bzw. 14 mg möglich, danach Dosisunterbrechung.</li> <li>• Dosisreescalation möglich.</li> <li>• Spezifische Regelungen für Dosisanpassungen aufgrund von UE.</li> <li>• Abbruch der Studienmedikation im Fall von UE CTCAE-Grad 4 (ausgenommen vorübergehende Laborwertabweichungen und Neutropenien, die 7 oder weniger Tage dauern).</li> </ul>	<b>Placebo</b> Siehe Intervention, ohne den aktiven Wirkstoff.

Intervention	Kontrolle
<b>Nicht erlaubte Begleitmedikation</b> <ul style="list-style-type: none"> <li>• Beginn jeglicher Tumor-gerichteter Therapie gegen TGCT, inklusive Strahlentherapie und Operation.</li> <li>• Intraartikuläre Steroidinjektionen zur Behandlung von TGCT.</li> <li>• Paracetamol-Einnahme über 3 g am Tag.</li> <li>• Prophylaktischer Einsatz von myeloischen Wachstumsfaktoren.</li> </ul>	

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; TGCT: Tenosynovialer Riesenzelltumor; UE: Unerwünschtes Ereignis.

## 2.3 Endpunkte

In diesem Kapitel werden die Eignung der Endpunkte hinsichtlich Operationalisierung, Patientenrelevanz und Validität beurteilt. Dazu wurden das Herstellerdossier, der zugehörige Studienbericht, das Studienprotokoll, der SAP und die in diesen Quellen zitierte Literatur herangezogen. Ergänzend wurden eigene Recherchen durchgeführt. Alle Endpunkte der eingeschlossenen 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 MOTION*

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Todesfälle <sup>1)</sup>	Mortalität	Ja	Ja
Tumoransprechen <sup>2)</sup>	Morbidität	Ergänzend	Ergänzend
Schmerzen mittels BPI-SF		Ja	Ja
PROMIS PF		Ja	Ja
Bewegungsumfang		Ja	Ergänzend
Steifigkeit mittels NRS		Ja	Ja
PGIS		Ja	Ergänzend
PGIC		Ja	Ja <sup>3)</sup>
EQ-5D-VAS		Ja	Ja
Unerwünschte Ereignisse	Sicherheit	Ja	Ja

<sup>1)</sup> Todesfälle wurden im Rahmen der Sicherheit erfasst.

<sup>2)</sup> Primärer Endpunkt.

<sup>3)</sup> Die allgemeine Krankheitssymptomatik mittels PGIC wird in der Nutzenbewertung berücksichtigt. Die Items zu Auswirkungen des Tumors auf Bewegungsumfang und körperliche Funktionsfähigkeit werden im Anhang dargestellt.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; NRS: Numerische Bewertungsskala; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; pU: pharmazeutischer Unternehmer.

### 2.3.1 Mortalität

#### Todesfälle

Der Endpunkt „Todesfälle“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Todesfälle wurden in der Studie MOTION im Rahmen der Sicherheit als UE, die zum Tod führten, innerhalb der Sicherheitspopulation erfasst. Informationen über UE und SUE werden ab dem Datum der Unterzeichnung der Einverständniserklärung bis 30 Kalendertage nach der letzten Dosis des Studienmedikaments erhoben.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

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

#### Validität

Der Endpunkt wird als valide bewertet.

### 2.3.2 Morbidität

#### Körperliche Funktionsfähigkeit mittels PROMIS PF

Der Endpunkt „PROMIS PF“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Im Rahmen der Studie MOTION wurde die körperliche Funktionsfähigkeit anhand von 15 Fragen aus der standardisierten Itembank des „Patient-Reported Outcomes Measurement Information System Physical Function“ (PROMIS PF) bewertet. Abhängig von der Lokalisation des Tumors werden daraus 2 Skalen gebildet: Die erste Skala umfasst 13 Fragen und ist an Personen mit Tumoren in den unteren Extremitäten (z. B. Knie) gerichtet, die zweite Skala umfasst 11 Fragen und richtet sich an Personen mit Tumoren in den oberen Extremitäten (z. B. Finger). Dabei überschneiden sich 9 der Fragen, d. h. sie sind sowohl in der Skala für die unteren als auch in der Skala für die oberen Extremitäten enthalten. Die 15 Items des Instruments werden unabhängig von der Lokalisation des Tumors abgefragt. Beantwortet wurden die Items elektronisch in einer Applikation auf Likert-Skalen mit jeweils 5 Punkten. Die Fragen beziehen sich auf den gegenwärtigen Moment der Beantwortung und haben keinen spezifischen Bezugszeitraum.

Zur Auswertung des PROMIS PF wurde ein T-Score mittels der Item-Kalibrierung des Assessment Center „Application Programming Interface“ (API) generiert, welches den Rohsummenwert in einen standardisierten Wert mit einem Mittelwert von 50 und einer Standardabweichung von 10 auf Grundlage der US-amerikanischen Gesamtbevölkerung skalierte. Ein T-Score wurde auf Basis des „API Scoring Manual“ berechnet und für die nachfolgenden Analysen verwendet, wenn eine Person mindestens die Hälfte der Fragen zu einer Tumorlokalisation beantwortet hatte. Fragen mit fehlenden Antworten wurden bei der Auswertung nicht berücksichtigt.

Der pU legte post hoc eine Responseschwelle für den PROMIS PF von mindestens 8,025 Punkten gegenüber dem Ausgangswert fest.

### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar. Fragen mit fehlenden Antworten wurden bei der Auswertung nicht berücksichtigt, daher ist davon auszugehen, dass der pU das von der PROMIS-Gruppe empfohlene „Response Scoring Pattern“ nicht eingesetzt hat.

### Patientenrelevanz

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

### Validität

Die Selektion der Items für die vorliegenden PROMIS-Messinstrumente erfolgte über Literaturreviews, klinische Expertise sowie Patienteninterviews [11]. Die Relevanz und Inhaltsvalidität von Items der Itembank des PROMIS PF wurde zunächst in halbstandardisierten Interviews mit 22 Patientinnen und Patienten mit TGCT in unterschiedlichen Gelenken und Körperregionen untersucht. Aus den 121 Items der Itembank des PROMIS PF wurden zunächst diejenigen extrahiert, die von  $\geq 20\%$  der Patientinnen und Patienten als relevant bewertet wurden. Bei der weiteren Selektion und Reduktion auf 15 Items wurden weitere Kriterien, wie die Spezifität der Items, Charakteristika der Items (auf Basis der „Item Response Theory“), die Häufigkeit der jeweiligen Auswirkung/Aktivität sowie Informationen aus den qualitativen Interviews zur Konzeptbeschreibung („Concept elicitation“) berücksichtigt [12]. Von den 15 Items wurden 9 sowohl von Betroffenen mit Tumoren der oberen und unteren Extremitäten bestätigt, 4 Items ausschließlich von Betroffenen mit Tumoren der unteren Extremitäten und 2 Items ausschließlich von Betroffenen mit Tumoren der oberen Extremitäten. In die Studie waren jedoch lediglich 2 Personen mit Tumoren der oberen Extremitäten eingeschlossen und die Aussagekraft dieser Ergebnisse ist eingeschränkt.

Anschließend erfolgte die weitere Validierung des Messinstruments anhand von Daten von 120 Personen mit nicht-resektablem TGCT aus einer randomisierten Interventionsstudie, davon jedoch lediglich  $n = 10$  mit TGCT in den oberen Extremitäten [19]. In der Folge konnte über eine konfirmatorische Faktorenanalyse lediglich die Eindimensionalität der Items für die unteren Extremitäten geprüft werden, welche auf Basis der gezeigten Ergebnisse auch nicht uneingeschränkt bestätigt werden konnten. Dagegen zeigte sich eine gute Interne Konsistenz sowie Test-Retest-Reliabilität für beide Varianten des Fragebogens. Korrelationen zu konstrukt-nahen Messinstrumenten, wie EQ-5D-VAS, BPI-SF und NRS-Steifigkeit, waren erwartungsgemäß moderat bis hoch ausgeprägt. Zuletzt ergeben sich erste Hinweise für die Änderungssensitivität über Auswertungen zu longitudinalen Veränderungen im Studienverlauf gemessen anhand der globalen Fragen zur physischen Funktionalität und der allgemeinen Krankheitssymptomatik (PGIC), die aufgrund der geringen Patientenzahlen in den Analysen jedoch limitiert sind. Insgesamt erscheint der PROMIS PF für die unteren Extremitäten ausreichend valide, für die oberen Extremitäten ist der PROMIS PF aufgrund der geringen Patientenzahlen in den Validierungsstudien von unklarer Validität.

Kritisch ist zu sehen, dass Fragen mit fehlenden Antworten nicht berücksichtigt wurden. Es ist unklar, wie vollständig die Items jeweils ausgefüllt wurden; zur Auswertung mussten jedoch mindestens 50 % der Items der entsprechenden Version beantwortet worden sein. Gemäß PROMIS-Manual sollte die Nutzung der „Response Scoring Pattern“ aufgrund des besseren Umgangs mit fehlenden Werten bevorzugt werden [13]. Dies war offensichtlich nicht der Fall und aus den Unterlagen geht die Transformierungsmethode, die für die Auswertungen verwendet wurde, nicht eindeutig hervor.

Zur Begründung der Responseschwelle von 8,025 Punkten referenziert der pU auf ein Nutzenbewertungsverfahren im Anwendungsgebiet „Morbus Pompe“, bei dem dieser Schwellenwert für den „PROMIS PF Short Form 20a“ angewendet und akzeptiert wurde [14]. Die 8,025 Punkte stellen dabei 15 % der Skalenspannweite der Konversionstabellen aus dem Manual des PROMIS PF



spezifisch für den „PROMIS PF Short Form 20a“ dar. Konversionstabellen für die im vorliegenden Nutzenbewertungsverfahren eingesetzten Messinstrumente konnten nicht identifiziert werden. Aus diesem Grund wird die Responseschwelle von 8,025 Punkten als nicht adäquat eingeschätzt und es werden in der vorliegenden Nutzenbewertung ausschließlich die stetigen Auswertungen präsentiert.

### **Bewegungsumfang**

Der Endpunkt „Bewegungsumfang“ wird in der Nutzenbewertung aufgrund unklarer Patientenrelevanz in der vorliegenden Operationalisierung lediglich im Anhang berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der Bewegungsumfang (ROM) des betroffenen Gelenks wurde mittels Goniometrie beurteilt; folgende Werte wurden erhoben:

- Aktive Bewegung: eigenständige, nicht assistierte Bewegung durch Studienteilnehmenden
- Passive Bewegung: Bewegung durch externes Einwirken der untersuchenden Person
- Aktive und passive Bewegung des nicht betroffenen, gegenüberliegenden Gelenks zum Vergleich (ausschließlich zu Screening und Ende von Studienteil 1).

Der Messwert des betroffenen Gelenks (in Grad) wurde verwendet, um einen relativen Bewegungsumfang abzuleiten, der durch Normalisierung der Messung anhand eines von der „American Medical Association“ (AMA) bereitgestellten Referenzstandardwerts pro Bewegung ermittelt wurde (aktiv sowie passiv).

Adäquat trainiertes, verblindetes Studienpersonal (Physiotherapeuten oder orthopädische Chirurgen) sollten die Bewertung des Bewegungsumfangs durchführen, die Dokumentation erfolgte auf einem speziellen Formular. Nach Möglichkeit sollte immer dieselbe bewertende Person die Untersuchung des Bewegungsumfangs bei einer Testperson durchführen.

Zum Screening wurde die Bewegung mit dem kleinsten relativen ROM-Wert (schlechtester) identifiziert, um darauffolgend die Veränderungen im relativen Bewegungsumfang bei dieser Bewegung zu bewerten. Falls es Gleichstände gibt, wird ein Durchschnitt dieser ROM-Werte bei jeder nachfolgenden Visite gebildet und als relativer ROM-Wert in der entsprechenden Visite berücksichtigt.

Zusätzliche Auswertungen waren für die passiven Messungen sowie die normalisierten aktiven und passiven Ergebnisse im Vergleich zum gegenüberliegenden, nicht-betroffenen Gelenk geplant.

##### *Bewertung*

Die Operationalisierung ist nur bedingt nachvollziehbar. Daten zu den Referenzstandards der AMA liegen nicht vor und es ist unklar, ob es für die bewertenden Personen ein Manual zur Testdurchführung gab. In den Studienunterlagen ist ein Training für die bewertenden Personen erwähnt, aber Informationen zu Inhalt oder Umfang liegen nicht vor.

#### Patientenrelevanz

Von den Patientinnen und Patienten wahrgenommene Einschränkungen im Bewegungsumfang von Gelenken können abhängig von der Operationalisierung patientenrelevant sein. Vorliegend erfolgte die Erhebung objektiv mittels Goniometrie, wobei nicht die absolute Beweglichkeit in Grad, sondern die relative Beweglichkeit im Vergleich zum Referenzstandard der AMA ausgewertet wurde. Es ist unklar, ob die mittels Goniometrie gemessenen Einschränkungen bzw. Veränderungen für Patientinnen und Patienten direkt spürbar sind. Weiterhin ist fraglich, ob die relativen Veränderungen zwischen unterschiedlichen Gelenken (z. B. Hüfte vs. Knie) vergleichbar sind.



In Modul 4 begründet der pU die Patientenrelevanz des Endpunkts mit dem direkten Zusammenhang zwischen dem Bewegungsumfang und den Auswirkungen auf die Aktivitäten des täglichen Lebens und führt als Beispiel diverse Aktivitäten auf, die ab bestimmten Bewegungseinschränkungen nicht mehr durchgeführt werden können. Die vom pU zitierte Literatur ist jedoch nicht ausreichend, um den Bewegungsumfang als Surrogat für Aktivitäten des täglichen Lebens anzuerkennen. Zudem sind tägliche Aktivitäten auch Bestandteil des PROMIS PF.

Insgesamt wird die Patientenrelevanz des Endpunkts „Bewegungsumfang“ in der vorliegenden Operationalisierung als unklar bewertet.

#### Validität

Die Erhebung des Bewegungsumfangs mittels Goniometrie sollte von adäquat trainiertem, verblindeten Studienpersonal (Physiotherapeuten oder orthopädische Chirurgen) durchgeführt werden, wobei die Bewertung möglichst immer von derselben Person übernommen werden sollte. Zudem liegen klare Regeln vor, welche Gelenkbewegung für die Bewertung berücksichtigt wird (schlechtester ROM-Wert) und wie mit Gleichständen umgegangen wird. Auf Basis der vorliegenden Informationen erscheint eine objektive und valide Erhebung des Endpunkts möglich. Allerdings fehlen Validierungsdaten hinsichtlich Reliabilität und Validität, um eine abschließende Bewertung vorzunehmen. So korrelierte in einer Validierungsstudie zur NRS-Steifigkeit (siehe unten) die von den Patientinnen und Patienten wahrgenommene Steifigkeit nur mäßig ( $r = -0,31$ ) mit einer passiven ROM-Bewertung [19].

#### **Steifigkeit mittels NRS**

Der Endpunkt „Steifigkeit mittels NRS“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Die Steifigkeit des betroffenen Gelenks wurde im Rahmen der Studie MOTION anhand einer numerischen Bewertungsskala (NRS) für die stärkste Steifigkeit bestimmt. Es handelt sich um eine einzelne Frage, bei welcher die Patientinnen und Patienten aufgefordert werden, die stärkste Steifigkeit des betroffenen Gelenks in den letzten 24 Stunden zu bewerten. Dabei wird eine Skala von 0 bis 10 verwendet, wobei 0 „keine Steifigkeit“ und 10 „die stärkste vorstellbare Steifigkeit“ bedeutet.

Baseline ist definiert als der Durchschnitt aller Erhebungen, die innerhalb von 14 Tagen vor Behandlungsbeginn erfolgt sind. Dabei mussten mindestens 4 Messungen verfügbar sein. Zum Ende eines jeden Zyklus wird ein Wert für die NRS aus dem Durchschnitt der Erhebungen der Wochen 2 und 3 des Zyklus gebildet (z. B. für Woche 25 die Werte aus Woche 2 und 3 des 6. Zyklus). In diesen Zeiträumen wurde die NRS jeden zweiten Tag abgefragt. Auch hier mussten mindestens 4 Messungen verfügbar sein.

Geplant waren eine Auswertung der Mittelwertdifferenz zu Woche 25 sowie Responderanalysen für eine Verbesserung mit einem Responsekriterium vom -2 Punkten.

##### *Bewertung*

Die Operationalisierung ist weitestgehend nachvollziehbar. Unklar ist, weshalb die NRS nicht analog zu den Items des BPI-SF täglich erhoben wurde, sondern nur an jedem zweiten Tag und zudem nicht in Woche 4 eines jeden Behandlungszyklus (siehe Tabelle 7).

### Patientenrelevanz

Gelenksteifigkeit ist ein belastendes Symptom von TGCT, das von Patientinnen und Patienten häufig berichtet wird [1,11] und zusammen mit dem Symptom „Schmerzen“ ein Einschlusskriterium für die Studie MOTION war. Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevant entsprechend § 2 Satz 3 AM-NutzenV angesehen.

### Validität

In einer qualitativen Studie gaben alle 22 Personen mit TGCT an, dass die NRS-Steifigkeit verständlich und relevant ist [11]. Eine umfangreichere Validierung des Messinstruments erfolgte anhand von Daten von 120 Personen mit nicht-resektablem TGCT aus einer randomisierten Interventionsstudie [19]. Die Test-Retest-Reliabilität für einen Vergleich wöchentlicher Mittelwerte der NRS war hoch und zu Baseline zeigten sich sehr hohe Korrelationen zum Item des BPI-SF ( $r = 0,83$ ), jedoch nur geringe Korrelationen zur EQ-5D-VAS ( $r = -0,21$ ) und einer passiven ROM-Bewertung ( $r = -0,31$ ). Hinsichtlich der Änderungssensitivität zeigten sich deutliche Verringerungen in der NRS-Steifigkeit für Personen, die sich im Außenkriterium PGIC-Steifigkeit verbessert haben. Diese waren deutlich stärker ausgeprägt im Vergleich zu Personen mit keiner Veränderung oder einer Verschlechterung. Aufgrund geringer Fallzahlen sind die Analysen jedoch limitiert. Weitere Einschränkungen ergeben sich durch die unterschiedliche Operationalisierung der NRS-Steifigkeit in der Validierungsstudie (tägliche Erhebung i. V. m. wöchentlichen Auswertungen) im Vergleich zur Studie MOTION (Erhebung alle 2 Tage i. V. m. 2-wöchentlichen Auswertungen), weshalb unklar ist, ob sich die Validierungsergebnisse übertragen lassen. Insgesamt liegt keine Begründung vor, weshalb die NRS-Steifigkeit in der Behandlungsphase nicht täglich erhoben wurde, sondern nur jeden zweiten Tag in den ersten 3 Wochen eines Behandlungszyklus. Da sich zu den jeweiligen Auswertungszeitpunkten immer auf Woche 2 und 3 eines jeweiligen Behandlungszyklus bezogen wird und diese verglichen werden mit der 2-wöchigen Erhebung zu Baseline, liegt beiden Erhebungen eine unterschiedliche Anzahl an NRS-Werten zugrunde. Da die NRS-Steifigkeit jedoch ein Einzelitem mit einer einfachen NRS-Skala erhebt, ist dennoch von der Validität des Instruments für Personen mit TGCT auszugehen.

Es liegen darüber hinaus keine Informationen zur Vollständigkeit der Tagebucheinträge für die NRS vor und es ist unklar, wie hoch der Anteil fehlender Werte in den Analysen tatsächlich ist. Um zu den entsprechenden Zyklen einen Wert zu bilden, waren 4 verfügbare von 7 möglichen Messungen ausreichend. Weiterhin ist zu berücksichtigen, dass keine Erhebungen zu Woche 4 eines jeden Zyklus vorgesehen waren und sich die Ergebnisse für die NRS-Steifigkeit auf einen Zeitraum beziehen, der der angegebenen Analysevisite (z. B. Woche 21 – Zyklus 6 Tag 1) 2 bis 3 Wochen vorausgeht.

Die a priori definierte Responseschwelle für eine Verbesserung von -2 Punkten entspricht einer Veränderung von mindestens 15 % der Skalenspannweite der NRS und wird für die vorliegende Nutzenbewertung berücksichtigt.

### **Schmerzen mittels BPI-SF**

Der Endpunkt „Schmerzen mittels BPI-SF“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Beim „Brief Pain Inventory – Short Form“ (BPI-SF) handelt es sich um einen Fragebogen bestehend aus 15 Items zur Erfassung der Schwere und Auswirkungen von Schmerzen. In der Studie MOTION wurden 2 dieser Fragen („Stärkster Schmerz“ und „Durchschnittlicher Schmerz“ in den letzten 24 Stunden) als einzelne Items erhoben und ausgewertet. Bei beiden Items bewerten die Personen ihre Schmerzen auf einer Skala von 0 bis 10 Punkten, wobei 0 „keine Schmerzen“ und 10 „schlimmste vorstellbare Schmerzen“ bedeutet.

Analog zur NRS-Steifigkeit ist Baseline definiert als der Durchschnitt aller Erhebungen, die innerhalb von 14 Tagen vor Behandlungsbeginn erfolgt sind; es mussten mindestens 4 Messungen verfügbar sein. Zum Ende eines jeden Zyklus wird ein Wert für den BPI-SF aus dem Durchschnitt der Erhebungen der Wochen 2 und 3 des Zyklus gebildet (z. B. für Woche 25 die Werte aus Woche 2 und 3 des 6. Zyklus). Auch hier mussten mindestens 4 Messungen verfügbar sein.

Die Einnahme narkotischer Analgetika wird von Screening bis Behandlungsende in einem Patiententagebuch erfasst (Medikation, Datum, Zeit, Anzahl der Tabletten). Bei Personen, die verschiedene Arten von narkotischen Analgetika und/oder Dosierungen verwenden, wird der Analgetika-Verbrauch berechnet, indem die äquianalgetische Dosierung in Morphinäquivalent-Dosen umgerechnet wird. Die Umrechnungsfaktoren wurden, soweit verfügbar, den Leitlinien der „World Health Organization“ (WHO) [23] und ansonsten den Leitlinien der „Centers for Disease Control and Prevention“ (CDC) [3] entnommen. In Fällen, in denen weder die WHO- noch die CDC-Leitlinien einen Umrechnungsfaktor für Morphinäquivalente enthielten, wurden alternative Leitlinien wie CONSORT (Consolidated Standards of Reporting Trials) [22] herangezogen.

Die Auswertung in den Studienunterlagen fokussierte auf „Stärkster Schmerz“. Aber auch das Item „Durchschnittlicher Schmerz“ wird, analog zu „Stärkster Schmerz“, ausgewertet.

Die Auswertung erfolgte über Responderanalysen unter Berücksichtigung der Schmerzmitteleinnahme, wobei Responder definiert waren als Personen mit mindestens 30 % Verbesserung im mittleren Wert der NRS im Vergleich zu Baseline für „Stärkster Schmerz“ und „Durchschnittlicher Schmerz“ des BPI-SF, ohne eine Zunahme der Einnahme narkotischer Analgetika um 30 % oder mehr. Für die Schmerzmitteleinnahme wurden dabei jeweils die gleichen Erhebungszeiträume betrachtet. Diese Analyse war prädefiniert.

Als Sensitivitätsanalyse waren Responder definiert als Personen mit mindestens 30 % Verbesserung im mittleren Wert der NRS für „Stärkster Schmerz“ und „Durchschnittlicher Schmerz“ des BPI-SF und ohne irgendeine Zunahme der Einnahme narkotischer Analgetika.

In Modul 4 präsentiert der pU post hoc zusätzlich stetige Auswertungen für die Veränderung der Items „Stärkster Schmerz“ und „Durchschnittlicher Schmerz“ gegenüber Baseline sowie eine Responderanalyse unabhängig von der Einnahme narkotischer Analgetika. Eine Responderanalyse mit dem Responsekriterium von 15 % der Skalenspannweite wurde nicht vorgelegt.

### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

Schmerzen sind ein belastendes Symptom von TGCT, das von Patientinnen und Patienten häufig berichtet wird [1,11] und zusammen mit Gelenksteifheit ein Einschlusskriterium für die Studie MOTION war. Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevant entsprechend § 2 Satz 3 AM-NutzenV angesehen.

### Validität

Die Validität des BPI-SF wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im zugrundeliegenden Anwendungsgebiet als valide eingeschätzt.

Analog zum Endpunkt „Steifigkeit mittels NRS“ ist unklar, wie hoch der Anteil fehlender Werte in den Analysen ist, da keine Informationen zur Vollständigkeit der Tagebucheinträge für den BPI-SF vorliegen. Um zu den entsprechenden Zyklen einen Wert zu bilden, waren hier lediglich 4 verfügbare von 14 möglichen Messungen ausreichend. Auch beim BPI-SF waren keine Erhebungen zu Woche 4 eines jeden Zyklus vorgesehen und die Ergebnisse für die Visiten beziehen sich auf 2 bis 3 Wochen vorausgehende Werte.

Hinsichtlich der Schmerzmitteleinnahme wurden ausschließlich narkotische Analgetika betrachtet. Es konnte keine Begründung identifiziert werden, weshalb nicht auch nicht-narkotische Analgetika

(Paracetamol, Metamizol oder nichtsteroidale Antirheumatika) berücksichtigt wurden. Allgemein liegen keine Informationen zu Dosierung und Dauer analgetischer Komedikationen während der Studie vor. Zwar sind Begleitmedikationen dokumentiert, jedoch bilden diese Veränderungen im Studienverlauf nicht ab.

Die prädefinierte Responseschwelle von 30 % im Vergleich zu Baseline bezieht sich auf den individuellen Baseline-Wert jeder teilnehmenden Person und nicht auf die Skalenspannweite. Um in jedem Fall eine Verbesserung erreichen zu können, die der im IQWiG-Methodenpapier beschriebenen Responseschwelle von 15 % der Skalenspannweite entspricht, sind Werte zu Baseline von mindestens 5 Punkten erforderlich, damit eine Veränderung um 30 % einer Veränderung von mindestens 15 % der Skalenspannweite entspricht. Anhand der Werte zu Baseline ist ersichtlich, dass dies für einen relevanten Anteil der Patientinnen und Patienten nicht zutrifft. Verteilungswerte zur genaueren Quantifizierung dieser Patientengruppe fehlen jedoch.

Aufgrund dieser Limitationen werden im Dossier ausschließlich die stetigen Auswertungen ohne Einbezug von Daten zu narkotischen Analgetika für die beiden Items des BPI-SF berücksichtigt. Aus den Studiendaten geht hervor, dass lediglich 2 Personen aus jedem Studienarm überhaupt eine Erhöhung narkotischer Analgetika hatten und keine Person über 30 %, sodass keine Verzerrung durch Erhöhung narkotischer Analgetika zu erwarten ist.

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

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

#### Operationalisierung

##### *Beschreibung*

Der allgemeine Gesundheitszustand wurde über die visuelle Analogskala des European Quality of Life 5-Dimension (EQ-5D-VAS) erhoben. Die EQ-5D-VAS ist eine Skala von 0 bis 100, auf der Patientinnen und Patienten ihren Gesundheitszustand einschätzen. Ein Wert von „0“ entspricht dem denkbar schlechtesten Gesundheitszustand, ein Wert von „100“ dem denkbar besten Gesundheitszustand.

Es werden sowohl die Veränderung gegenüber Baseline als auch Responderanalysen zur Verbesserung um  $\geq 15$  % der Skalenspannweite zu Woche 25 berichtet.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

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

#### Validität

Die Validität der EQ-5D-VAS wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im zugrundeliegenden Anwendungsgebiet als valide eingeschätzt. Berücksichtigt werden Responderauswertungen mit einer Verbesserung um  $\geq 15$  % der Skalenspannweite.

## **Körperliche Funktionsfähigkeit / Bewegungsumfang mittels PGIS**

Der Endpunkt „PGIS“ wird in der Nutzenbewertung aufgrund unklarer Validität und Mehrfacherfassung ergänzend dargestellt.

### Operationalisierung

#### *Beschreibung*

Der „Patient Global Impression of Severity“ (PGIS) ist ein selbstberichteter Fragebogen, bestehend aus 2 Items zur Erfassung des Schweregrads der Auswirkungen des Tumors zum einen auf den Bewegungsumfang (PGIS-ROM, Item 1) und zum anderen auf die körperliche Funktionsfähigkeit (PGIS-PF, Item 2) am Tag der Befragung. Die Bewertung erfolgte auf einer 5-stufigen Skala:

- keine Symptome
- mild
- moderat
- schwer
- sehr schwer

Ausgewertet wurde die Veränderung der PGIS-Werte im Vergleich zu Baseline sowie der Anteil der Personen mit einer Reduktion (Verbesserung) im PGIS um mindestens 1 Punkt gegenüber Baseline für jedes Item separat.

#### *Bewertung*

Die Operationalisierung ist weitgehend nachvollziehbar.

### Patientenrelevanz

Die Einschätzung der Patientinnen und Patienten zur physischen Funktionalität sowie zum Bewegungsumfang betroffener Gelenke wird prinzipiell als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

### Validität

Die Validität des Endpunkts „PGIS“ hinsichtlich der allgemeinen Symptome wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und als valide eingeschätzt. Vorliegend wird mit beiden Items zum PGIS jedoch nicht die allgemeine Symptomatik, sondern die Auswirkung des Tumors zum einen auf den Bewegungsumfang (Item 1) und zum anderen auf die körperliche Funktionsfähigkeit (Item 2) erfasst. Ziel der Erhebung war zudem die Validierung der Endpunkte „Bewegungsumfang“ und „PROMIS PF“ im Rahmen der Studie MOTION, weshalb die Responderanalysen post hoc für die vorliegende Nutzenbewertung durchgeführt wurden. Hinsichtlich der Validität der Items „PGIS-ROM“ und „PGIS-PF“ liegen keine psychometrischen Untersuchungen vor. Es ist unklar, inwiefern mit Einzelitems mit jeweils 5 Ausprägungen die physische Funktionalität bzw. der Bewegungsumfang des betroffenen Gelenks adäquat erfasst werden kann.

Da mit dem PROMIS PF bereits ein weiteres, valides Instrument zur Erfassung der körperlichen Funktionalität in der Studie MOTION zum Einsatz kam, werden die Ergebnisse zum PGIS-PF lediglich ergänzend zum PROMIS PF berücksichtigt. Patientenberichtete Einschränkungen im Bewegungsumfang werden bereits über den Endpunkt „Steifigkeit mittels NRS“ erfasst. Aufgrund der Doppelerfassung erfolgt für die Ergebnisse zum PGIS-ROM ebenfalls eine ergänzende Darstellung.

### **Allgemeine Krankheitssymptomatik mittels PGIC**

Beim Endpunkt „PGIC“ wird Item 3 zur allgemeinen Symptomatik in der Nutzenbewertung berücksichtigt. Die Items 1 (PGIC-ROM) und 2 (PGIC-PF) werden aufgrund unklarer Validität und Mehrfacherfassung aus Transparenzgründen im Anhang dargestellt.

#### Operationalisierung

##### *Beschreibung*

Der „Patient Global Impression of Change“ (PGIC) ist ein selbstberichteter Fragebogen, bestehend aus 3 Items zur Erfassung der Veränderung der Auswirkungen des Tumors auf den Bewegungsumfang (PGIS-ROM, Item 1), auf die körperliche Funktionsfähigkeit (PGIS-PF, Item 2) und auf die allgemeinen Symptome an der Tumorstelle (PGIC-Krankheitssymptomatik, Item 3) zum jeweiligen Erhebungszeitpunkt im Vergleich zu Baseline. Die Bewertung erfolgte auf einer 7-stufigen Skala:

- sehr viel besser
- viel besser
- wenig besser
- keine Änderung
- wenig schlechter
- Viel schlechter
- sehr viel schlechter

Gemäß SAP sollte der Anteil der Personen mit „viel besser“ oder „sehr viel besser“ für jedes Item separat ausgewertet werden. Dagegen wird in Modul 4 eine Post-hoc-Analyse berichtet, in der der Anteil der Personen mit einer Verbesserung im PGIC („wenig besser“, „viel besser“ oder „sehr viel besser“) ausgewertet wird.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz und Validität

Die Einschätzung der Patientinnen und Patienten zur Veränderung der jeweils eigenen erkrankungsspezifischen Symptomatik ist grundsätzlich als patientenrelevant zu bewerten. Methodisch werden dabei i. d. R. Instrumente bevorzugt, die sowohl zu Baseline- als auch zu Post-Baseline-Studienvisiten erhoben werden. Veränderungseinschätzungen gegenüber Baseline ohne eigenständige Baseline-Erhebung können demgegenüber unter Umständen weniger aussagekräftige Ergebnisse liefern. Es gibt Hinweise darauf, dass bei Einschätzungen von Veränderungen größtenteils der gegenwärtige Gesundheitszustand erfasst wird. Dieser Effekt wird umso größer, je länger das Erinnerungsintervall ist [15,16,17]. Für die Visite zum Ende von Teil 1 der Studie MOTION müssen die Patientinnen und Patienten 24 Wochen zurückblicken, um sich an ihren Zustand zu Beginn der Studienteilnahme zu erinnern.

In der Studie MOTION wurde der PGIC in einer Version bestehend aus 3 getrennt ausgewerteten Items eingesetzt. Dabei entspricht das Item zur allgemeinen Symptomatik der regelhaften Operationalisierung globaler Veränderungen in klinischen Studien. Die beiden weiteren Items des PGIC erfragen Veränderungen bei den Auswirkungen des Tumors zum einen auf den Bewegungsumfang und zum anderen auf die körperliche Funktionsfähigkeit. Analog zum Endpunkt „PGIS“ ist unklar, inwiefern Einzelitems die physische Funktionalität bzw. den Bewegungsumfang des betroffenen Gelenks adäquat abbilden können. Zudem werden beide Konzepte in der Studie MOTION bereits mit anderen Erhebungsinstrumenten (PROMIS PF, NRS-Steifigkeit) erfasst. Für die Items liegen darüber hinaus vergleichbare Ergebnisse für den PGIS vor.

Zur Vermeidung von Doppelerfassung und aufgrund der Annahme, dass der PGIC im Vergleich zum PGIS keine wesentlichen zusätzlichen Informationen generiert, werden der PGIC-ROM und der PGIC-PF lediglich aus Transparenzgründen im Anhang dargestellt. Für das Item zur allgemeinen



Symptomatik liegen weder ähnliche Erhebungsinstrumente noch Daten aus dem PGIS vor und die Ergebnisse werden mit Hinweis auf den möglichen „Recall Bias“ für die Nutzenbewertung berücksichtigt.

Gemäß SAP waren Responder im PGIC definiert als Personen die „sehr viel besser“ und „viel besser“ angaben, wohingegen post hoc in Modul 4 zusätzlich auch solche Personen mit „wenig besser“ als Responder gewertet wurden. Zur Begründung verweist der pU auf das Nutzenbewertungsverfahren zu Mavacamten, in dem eine ähnliche Responderdefinition akzeptiert wurde. Prinzipiell werden a priori definierte Responderkriterien bevorzugt. Da die Anzahl der Responder zwischen den Behandlungsarmen in beiden Konstellationen ähnlich ausfällt, werden die Post-hoc-Responderanalysen für den PGIC als adäquat eingestuft.

## **Tumoransprechen**

Das Tumoransprechen wird in der vorliegenden Operationalisierung als nicht patientenrelevant bewertet. Da es sich in der Operationalisierung „ORR zu Woche 25“ um den primären Endpunkt der Studie MOTION handelt, werden die Ergebnisse aus Transparenzgründen im Anhang dargestellt.

### Operationalisierung

#### *Beschreibung*

Das Tumoransprechen bzw. die „Objektive Ansprechrates“ (ORR) ist definiert als der Anteil der Studienteilnehmenden mit einem kompletten Ansprechen (CR) oder partiellen Ansprechen (PR) zu Woche 25. Die Bestimmung des Responderstatus zu Woche 25 erfolgt gemäß Tabelle 25 im Anhang. Jegliche ungeplante Studienvsiste vor Woche 25 geht dabei analog zu Woche 13 in die Bestimmung des Responderstatus ein. Falls keine Bewertung zu Woche 25 erfolgte, gilt die Person als Non-Responder.

Zur Bewertung des Ansprechens wurde eine Magnetresonanztomographie (MRT) des betroffenen Gelenks zu Woche 13 und zu Woche 25 durchgeführt. Sofern das Prüfpersonal den Verdacht einer Krankheitsprogression hat, sollte zusätzliche ein MRT-Bild sofort zur Bewertung an den zentralen Anbieter geschickt werden. Die Beurteilung des Tumoransprechens erfolgte mittels unabhängiger radiologischer Überprüfung (IRR) durch zwei unabhängige Radiologen und gegebenenfalls durch eine Adjudikation (mRECIST und TVS werden vom zentralen Bildgebungsanbieter generiert). Im Fall der Bewertung gemäß RECIST erfolgte zusätzliche eine Beurteilung durch das Prüfpersonal. Zusätzlich zu den „Response Evaluation Criteria In Solid Tumors“ (RECIST, Version 1.1) wird das Ansprechen anhand folgender Kriterien bewertet:

- Tumolvolumen-Score (TVS): Anteil der Personen mit CR und PR zu Woche 25.
- mRECIST: Anteil der Personen mit CR und PR zu Woche 25.

#### **mRECIST**

Die modifizierten RECIST (mRECIST) sind identisch mit denen von RECIST, mit der Ausnahme, dass die Messungen entlang der kurzen Achse senkrecht zur längsten Abmessung und idealerweise auch an einer reproduzierbaren benachbarten Markierung, wie dem Oberschenkelknochen oder einer Sehne, vorgenommen werden. Die Summe der kurzaxialen Abmessungen der Zielläsionen wird für jeden Zeitpunkt berechnet und angegeben. Alle anderen Konventionen, die für RECIST gelten, gelten auch für mRECIST.

#### **TVS**

Der TVS ist ein semiquantitatives MRT-Bewertungssystem, das die Tumormasse beschreibt und auf 10%igen Schritten des geschätzten Volumens der maximal gedehnten Synovialhöhle oder Sehnenscheide basiert. Ein Tumor, dessen Volumen dem einer maximal gedehnten Synovialhöhle oder Sehnenscheide entspricht, wurde mit 10 bewertet; ein Score von 0 bedeutete, dass kein Tumor nachweisbar war.

Die Bestimmung des Ansprechens für jeden Zeitpunkt basiert auf der Kombination des Ansprechens für die Zielläsionen und dem Vorhandensein oder Nicht-Vorhandensein einer oder mehrerer neuer Läsionen. Zur Bestimmung des Tumoransprechstatus zum Ende von Studienteil 1 liegen genau definierte Regelungen unter Einbezug der vorherigen Erhebungen vor, sodass nur Personen mit Ansprechen zu Woche 25 als Responder gewertet werden.

#### Auswertungen

- Für den primären Endpunkt der Studie (ORR zu Woche 25) erfolgte die Bewertung nach RECIST gemäß IRR.
- ORR bis Woche 25 gemäß IRR, wobei alle Personen als Responder gewertet werden, die die Kriterien für ein Ansprechen mindestens einmal erfüllt haben.
- ORR zu Woche 25 nach mRECIST gemäß IRR (sekundär).
- ORR zu Woche 25 nach TVS gemäß IRR (sekundär).
- ORR zu Woche 25 nach RECIST gemäß Prüfpersonal.

#### Bewertung

Die Operationalisierung ist im Wesentlichen nachvollziehbar. Die zentrale Bewertung der radiologischen Aufnahmen durch eine IRR wird als adäquat eingeschätzt.

#### Patientenrelevanz

Die Bestimmung des Tumoransprechens erfolgte anhand RECIST [7] sowie mRECIST [2] und TVS [21]. Die Beurteilung beruht somit ausschließlich auf bildgebenden Verfahren und berücksichtigt keine von Patientinnen oder Patienten wahrnehmbare Symptomatik. Daher ist die Operationalisierungen des Endpunkts „Tumoransprechen“ nicht per se patientenrelevant. Die Verbesserung der Funktionseinschränkungen und der Symptomatik bei Schrumpfung des Tumors sollte sich zudem in anderen Endpunkten zeigen (z. B. PROMIS PF).

#### Validität

Die Erhebung und Bewertung der bildgebenden Verfahren anhand RECIST wird als valide eingeschätzt.

### **2.3.3 Lebensqualität**

Lebensqualität wurde in der Studie MOTION nicht erhoben.

### **2.3.4 Sicherheit**

#### **Unerwünschte Ereignisse**

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

#### Operationalisierung

##### *Beschreibung*

Ein unerwünschtes Ereignis (UE) ist als jegliches unerwünschtes medizinische Ereignis definiert, unabhängig vom kausalen Zusammenhang mit der Studienmedikation. Dies schließt alle neu auftretenden Ereignisse ein sowie bestehende Zustände, die sich in der Schwere oder Häufigkeit verschlechtern. UE wurden gemäß „Medical Dictionary for Regulatory Activities“ (MedDRA, Version 26.0 oder höher) nach Systemorganklasse (SOC) und Preferred Term (PT) kodiert. Symptome der untersuchten Krankheit dürfen nicht als UE erfasst werden, solange sie innerhalb der normalen täglichen Schwankungen oder des erwarteten Krankheitsverlaufs liegen. Dies schließt eine signifikante Verschlechterung ein, es sei denn die Verschlechterung war unerwartet und ist Teil der zu erhebenden Wirksamkeitsdaten.



### Schwerwiegendes unerwünschtes Ereignis

Ein UE wurde als schwerwiegendes unerwünschtes Ereignis (SUE) eingestuft, wenn es einem der folgenden Kriterien entsprach:

- Führt zum Tod.
- Ist lebensbedrohend (ein Ereignis, bei dem der Patient das Risiko hatte, während des Ereignisses bzw. der Reaktion zu versterben; hierzu gehören keine Ereignisse/Reaktionen, welche hypothetisch zum Tod führten könnten, wenn sie schwerer ausgeprägt wären).
- Erfordert eine stationäre Hospitalisierung oder die Verlängerung einer bereits bestehenden Hospitalisierung.
- Führt zu einer angeborenen Anomalie / einem Geburtsfehler.
- Ist ein wichtiges medizinisches Ereignis, das nicht zum Tod führt, lebensbedrohlich ist oder eine Hospitalisierung erforderlich macht, aber die Testperson gefährdet und einen medizinischen oder operativen Eingriff notwendig machen kann, um einen der zuvor genannten Punkte zu verhindern.

### Schwere UE

Der Schweregrad der UE wurde gemäß „Common Terminology Criteria for Adverse Events“ (CTCAE, Version 5.0) des National Cancer Institute (NCI) klassifiziert. Der Schweregrad eines UE, das nicht in der CTCAE-Skala aufgeführt ist, wurde wie folgt klassifiziert:

- Grad 1 (mild): Asymptomatische oder leichte Symptome; nur klinische oder diagnostische Beobachtungen; keine Intervention angezeigt.
- Grad 2 (moderat): Minimaler, lokaler oder nichtinvasiver Eingriff angezeigt; Einschränkung altersgemäßer instrumenteller Aktivitäten des täglichen Lebens. Kodiert nach MedDRA.
- Grad 3 (schwer): Schwerwiegend oder medizinisch bedeutsam, aber nicht unmittelbar lebensbedrohlich; Krankenhausaufenthalt oder Verlängerung des Krankenhausaufenthalts angezeigt; Behinderung; Einschränkung der Selbstversorgung bei Aktivitäten des täglichen Lebens.
- Grad 4 (lebensbedrohlich): Lebensbedrohliche Folgen; dringende Intervention angezeigt.
- Grad 5 (Tod): Tod im Zusammenhang mit UE.

Für die Auswertung werden alle behandlungsbedingten UE berücksichtigt; definiert als alle UE, die nach Erhalt der ersten Dosis der Studienmedikation bis 30 Tage nach letztmaligem Erhalt der Studienmedikation auftraten.

### *Bewertung*

Die Operationalisierung ist weitgehend nachvollziehbar. Inwiefern die 30-tägige Sicherheitsnachbeobachtung mit in die vorliegenden Ergebnisse zu den UE aus Teil 1 der Studie MOTION mit eingeht, ist offensichtlich abhängig von einer Teilnahme an Teil 2 der Studie. Der pU legt keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor und es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung mit in die Erfassung der UE eingingen. Allerdings durften Symptome der untersuchten Krankheit nicht als UE erfasst werden, solange sie innerhalb der normalen täglichen Schwankungen oder des erwarteten Krankheitsverlaufs liegen.

### Patientenrelevanz und Validität

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

### 2.3.5 Erhebungszeitpunkte

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

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

<div>Studienvisite</div> <div>Endpunkt</div>	Screening 14 Tage	Zyklus 1		Zyklus 2 bis 6		Ende Teil 1	EOT <sup>1)</sup> ± 7 Tage	FuP Sicherheit ± 5 Tage
		W1 T1	W3 T1 ± 1 Tag	W1 T1 ± 7 Tage	W2 T6	± 7 Tage		
Todesfälle	Kontinuierlich im Rahmen der Sicherheit							
Tumoransprechen <sup>2)</sup>	x			x <sup>3)</sup>		x	x	
Schmerzen mittels BPI-SF <sup>4)</sup>	x		x	x	x	x	x	
PROMIS PF <sup>5)</sup>	x <sup>6)</sup>		x	x	x	x	x	
Bewegungsumfang <sup>7)</sup>	x			x <sup>3)</sup>		x	x	
Steifigkeit mittels NRS <sup>8)</sup>	x		x	x	x	x	x	
PGIS <sup>5)</sup>	x <sup>6)</sup>		x	x	x <sup>9)</sup>	x	x	
PGIC <sup>5)7)</sup>			x	x	x <sup>9)</sup>	x	x	
EQ-5D-VAS <sup>5)</sup>	x <sup>6)</sup>		x	x		x	x	
Unerwünschte Ereignisse	Kontinuierlich							

<sup>1)</sup> Nicht erforderlich, falls die entsprechende Erhebung innerhalb der vergangenen 7 Tage durchgeführt wurde.

<sup>2)</sup> Endpunkt wird ergänzend dargestellt. Follow-up für das Tumoransprechen 12 Wochen nach letzter Dosis.

<sup>3)</sup> Nur an Tag 1 in Zyklus 4 und ausschließlich Bewertung des betroffenen Gelenks.

<sup>4)</sup> Tägliche Erhebung (Tagebuch), ausgenommen Woche 4 eines jeden Zyklus sowie zusätzlich Woche 1 Tag 1, Woche 2 Tag 7 und Woche 3 Tag 2 in Zyklus 1.

<sup>5)</sup> Zeitfenster von ± 2 Tagen während der Behandlungszyklen.

<sup>6)</sup> Erhebung an Tag 7 von Woche 2 des Screeningzeitraums. Zeitfenster von ± 7 Tagen für das Screening.

<sup>7)</sup> Endpunkt wird ergänzend bzw. aus Transparenzgründen im Anhang dargestellt.

<sup>8)</sup> Tägliche Erhebung im Screeningzeitraum. Erhebung jeden zweiten Tag im Behandlungszeitraum. Keine Erhebung in Woche 4 eines jeden Zyklus sowie zusätzlich Woche 1 Tag 1, Woche 2 Tag 7 und Woche 3 Tag 2 in Zyklus 1.

<sup>9)</sup> Nur PGIC-PF bzw. PGIS-PF.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; EOT: End of Treatment; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; FuP: Follow-up; NRS: Numerische Bewertungsskala; PF: Körperliche Funktionsfähigkeit; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; W: Woche; T: Tag.

## 2.4 Statistische Methoden

Die statistischen Analysen basieren auf dem SAP der Version 2.0 vom 31.05.2023. Zur Bestimmung von klinischen Relevanzschwellen für die NRS-Steifigkeit wurden qualitative Auswertungen anhand der klinischen Daten einer Subpopulation aus der Studie MOTION durchgeführt. Nach Angaben im Bericht wurden die Relevanzschwellen vor Entblindung bestimmt [9].

### Analysepopulationen

- Intention-to-Treat (ITT)-Population: Alle randomisierten Personen, nach zugewiesener Behandlung.
- Sicherheitspopulation: Alle Person, die mindestens eine Dosis der Studienmedikation erhalten haben; ausgewertet nach tatsächlich erhaltener Behandlung.

### Datenschnitte

Der vorliegende Datenschnitt vom 22.08.2023 entspricht dem Abschluss des vergleichenden Studienteil 1. Die Studie ist laufend. Interimanalysen waren gemäß Studienprotokoll nicht geplant.

### Präspezifizierte Subgruppenanalysen

In den Studienunterlagen waren für die Endpunkte „Tumorsprechen“, „Bewegungsumfang“, „PROMIS PF“, „Steifigkeit mittels NRS“, „EQ-5D-VAS“ und „BPI-SF“ folgende Subgruppenanalysen präspezifiziert:

- Geschlecht (weiblich vs. männlich)
- Alter ( $\geq 18$  bis  $< 50$  Jahre vs.  $\geq 50$  bis  $< 65$  Jahre vs.  $\geq 65$  Jahre)
- Subtyp der Krankheit (nodulär/lokalisiert vs. diffus)
- Lage des Tumors (untere Gliedmaßen vs. alle anderen Gliedmaßen)
- Erkrankungen an großen Gelenken (Schulter, Ellbogen, Hüfte oder Knie) vs. Erkrankung an kleinen Gelenken (andere Gelenke als Schulter, Ellbogen, Hüfte, Knie)
- Erkrankung im Bereich des Knies (ja vs. nein)
- Frühere Operationen (ja vs. nein)
- Vorherige Behandlung mit Imatinib oder Nilotinib (ja vs. nein)
- Standort in den USA (ja vs. nein)
- Standort in Europa (ja vs. nein)

Post hoc wurden vom pU für alle weiteren in Modul 4 präsentierten Analysen Subgruppenanalysen durchgeführt.

Aufgrund des geringen Stichprobenumfangs ( $N < 10$  in der Subgruppe „ $\geq 65$  Jahre“) werden die Subgruppenanalysen nach „Alter“ nicht berücksichtigt. Dabei wird dem pU gefolgt, der die Ergebnisse für diese Subgruppenanalysen in Modul 4 auch nicht berichtet.

In der Nutzenbewertung werden nur die Ergebnisse dargestellt, bei denen eine Effektmodifikation mit einer statistisch signifikanten Interaktion zwischen Behandlung und Subgruppenmerkmal ( $p < 0,05$ ) vorliegt. Zudem werden ausschließlich Subgruppenergebnisse dargestellt, bei denen mindestens in einer Subgruppe ein statistisch signifikanter und relevanter Effekt vorliegt.

## **Präspezifizierung geplanter und durchgeführter Analysen**

### Kontinuierliche Endpunkte

Gemäß SAP wurden „Mixed Models for Repeated Measures“ (MMRM) unter Verwendung des Sandwich-Schätzers zur Schätzung der Varianz-Kovarianz-Matrix durchgeführt. Statistische Vergleiche zwischen den Behandlungsgruppen werden auf der Grundlage einer Gegenüberstellung zu Woche 25 vorgenommen.

Post hoc wurde für das vorliegende Nutzendossier zusätzlich ein „Gesamteffekt“ für die Behandlung als MMRM sowie die Differenz des „Gesamteffekts“ zwischen den Behandlungsgruppen vom Ausgangswert zu Woche 25 beschrieben. Darüber hinaus wurde post hoc die Schätzung des Hedges' g der standardisierten mittleren Differenz für die Gesamtveränderung gegenüber dem Ausgangswert zu Woche 25 und das zugehörige zweiseitige 95%-Konfidenzintervall (KI) angegeben.

Für explorative Endpunkte waren keine inferenzstatistischen Testungen geplant.

### Binäre Endpunkte

Im Studienbericht war eine Auswertung binärer Endpunkte anhand eines Vergleichs der Responderanteile zwischen den beiden Behandlungsgruppen unter Verwendung eines zweiseitigen Cochran-Mantel-Haenszel-Tests geplant, stratifiziert nach den Stratifikationsfaktoren der Randomisierung.

Post hoc wurden für Modul 4 relative Risiken (RR) für binäre Endpunkte berechnet. Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Werts wurde eine log-binomiale Regression verwendet, wobei Placebo als Referenzgruppe diente. Die Schätzung des Behandlungseffekts basierte auf einem stratifizierten statistischen Modell.

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

Im Fall von Responderanalysen wurden fehlende Werte nach Baseline mittels Non-Responder-Imputation ersetzt.

### Sensitivitätsanalysen für fehlende Werte

Wenn in Woche 25 bei einem der wichtigsten sekundären Wirksamkeitsendpunkte „PROMIS PF“, „Bewegungsumfang“, „Steifigkeit mittels NRS“ und „EQ-5D-VAS“ mehr als 10 % der Daten fehlen, wird die Auswirkung unter anderem mittels folgender Sensitivitätsanalysen untersucht:

- Mehrfache Imputation nach randomisierter Behandlung, um zufällige fehlende Daten nachzubilden.
- Pattern Mixture Model (Behandlungsabbrecher aufgrund von UE werden behandelt, als ob sie Placebo erhalten haben)

## **Einschätzung der statistischen Auswertungen**

In Modul 4 präsentiert der pU für die stetigen Auswertungen post hoc berechnete MMRM-Analysen und p-Werte unter Bezugnahme auf den „Gesamteffekt“ der Behandlung. Die genaue Herangehensweise sowie die Definition des Begriffs „Gesamteffekt“ sind unklar und von Seiten des pU konnte keine Begründung für die Post-hoc-Auswertungsmethodik identifiziert werden. Die a priori geplanten MMRM-Auswertungen der Veränderung zu Woche 25 werden aufgrund der Präspezifizierung sowie den sich über den Zeitverlauf endpunktübergreifenden abzeichnenden stetigen Verbesserungen im Interventionsarm bevorzugt und zur Nutzenbewertung herangezogen.

Die post hoc durchgeführten Berechnungen der RR erscheinen prinzipiell adäquat. Für die Auswertungen der UE gibt der pU in Modul 4 jedoch an, dass die Analysen stratifiziert nach den

Stratifizierungsfaktoren „Tumorlokalisation“ und „Region“ erfolgten. Nachberechnungen sowie weitere Angaben aus den Unterlagen des pU deuten jedoch darauf hin, dass die präsentierten Schätzer unadjustiert für die Stratifizierungsfaktoren sind.

Responderanalysen für Verschlechterungen wurden weder in den Studienunterlagen prädefiniert noch post hoc für das Modul 4 durchgeführt. Aufgrund des Therapiefokus auf Verbesserungen und des nur langsam progredienten Erkrankungsverlaufs, bei einer Beobachtungsdauer von 24 Wochen, sowie der Tatsache, dass während Teil 1 der Studie MOTION nur 1 Person eine Progression erlitt, erscheint der Fokus auf Verbesserungen plausibel.

Die im SAP geplanten Sensitivitätsanalysen für fehlende Werte bei wichtigen sekundären Endpunkten wurden nicht für die Post-hoc-Analysen durchgeführt. Relevante Sensitivitätsanalysen liegen somit nur für die Ergebnisse des PROMIS PF vor.

## 2.5 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 7: Verzerrungspotential der Studie MOTION

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
MOTION	Ja	Ja	Unklar <sup>1)</sup>	Unklar <sup>1)</sup>	Nein	Nein	Niedrig

<sup>1)</sup> Aufgrund spezifischer charakteristischer UE wie beispielsweise Periorbitalödem, Gesichtssödem, Pruritus und Ausschlag kann eine potentielle Entblindung nicht ausgeschlossen werden.

Abkürzungen: UE: Unerwünschte Ereignisse.

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

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

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
<b>Morbidität</b>					
PROMIS PF	Unklar <sup>1)</sup>	Nein <sup>2)</sup>	Nein	Nein	Hoch
Steifigkeit mittels NRS	Unklar <sup>1)</sup>	Nein <sup>2)3)</sup>	Nein	Nein	Hoch
Schmerzen mittels BPI-SF	Unklar <sup>1)</sup>	Nein <sup>2)</sup>	Nein	Nein	Hoch
EQ-5D-VAS	Unklar <sup>1)</sup>	Nein <sup>2)</sup>	Nein	Nein	Hoch
PGIS	Unklar <sup>1)</sup>	Nein <sup>2)</sup>	Nein	Nein	Hoch
PGIC	Unklar <sup>1)</sup>	Nein <sup>2)</sup>	Ja <sup>4)</sup>	Nein	Hoch
<b>Sicherheit</b>					
Schwere UE (Gesamtrate, SOC/PT)	Ja	Ja	Nein	Ja <sup>5)</sup>	Unklar <sup>6)</sup>
SUE (Gesamtrate, SOC/PT)	Ja	Ja	Nein	Ja <sup>5)</sup>	Unklar <sup>6)</sup>
Abbruch wegen UE (Gesamtrate)	Unklar <sup>1)</sup>	Ja	Nein	Ja <sup>5)</sup>	Hoch
UE, jeglichen Schweregrads (SOC/PT)	Unklar <sup>1)</sup>	Ja	Nein	Ja <sup>5)</sup>	Hoch

<sup>1)</sup> Aufgrund spezifischer charakteristischer UE wie beispielsweise Periorbitalödem, Gesichtsödem, Pruritus und Ausschlag kann eine potentielle Entblindung nicht ausgeschlossen werden.

<sup>2)</sup> Der Anteil fehlender Werte in beiden Studienarmen für die Auswertung zu Woche 25 liegt bei über 20%.

<sup>3)</sup> Unterschied im Anteil fehlender Werte (non-Responder Imputation) zwischen den Behandlungsarmen > 5% (75,9% im Vimseltinib Arm vs. 67,5% im Placebo-Arm).

<sup>4)</sup> Gemäß CSR wurden Personen mit Angabe „viel besser“ und „sehr viel besser“ als Responder gewertet. In den Post-hoc-Analysen für Modul 4 waren zusätzlich Personen mit „wenig besser“ als Responder definiert. Das daraus resultierende Verzerrungspotential wird jedoch als gering gewertet, da auch bei den Auswertungen im CSR deutliche Unterschiede in den Responderanteilen zwischen den Studienarmen ersichtlich sind.

<sup>5)</sup> Es ist davon auszugehen, dass die Analysen für die UE unstratifiziert erfolgten.

<sup>6)</sup> Aufgrund der Effektschätzer und der p-Werte ist trotz unklarem Verzerrungspotential die Ergebnissicherheit nicht eingeschränkt.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; CSR: Clinical Study Report; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ITT: Intention to Treat; NRS: Numerische Bewertungsskala; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; PT: Preferred Term; SOC: Systemorganklasse; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

### 3 Ergebnisse der eingeschlossenen Studie

#### 3.1 Studiencharakteristika und Studienmedikation

Von den 146 gescreenten Personen wurden 123 Personen in die Studie MOTION eingeschlossen und randomisiert. Alle randomisierten Personen erhielten ihre zugewiesene Behandlung, abgesehen von einer Person aus dem Placebo-Arm, die nicht für einen Studieneinschluss geeignet war, aber trotzdem verfrüht randomisiert jedoch nicht behandelt wurde. Die allgemeinen Angaben zur Studie MOTION sind in Tabelle 10 aufgeführt. Lediglich 1 Person (2,6 %) aus dem Placebo-Arm beendete die Studienmedikation frühzeitig aufgrund von Krankheitsprogression und wechselte in Studienteil 2 auf Vimseltinib.

*Tabelle 9: Allgemeine Angaben; Studie MOTION*

<b>Studie MOTION Allgemeine Angaben</b>	<b>Vimseltinib</b>	<b>Placebo</b>
ITT-Population <sup>1)</sup> , n (%)	83 (100)	40 (100)
Sicherheitspopulation, n (%)	83 (100)	39 (97,5) <sup>2)</sup>
Abbruch der Einnahme der Studienmedikation in Teil 1, n (%)	9 (10,8)	5 (12,5)
Aufgrund von:		
UE	4 (4,8)	0 (0)
Krankheitsprogression	0 (0)	1 (2,5)
Entscheidung Ärztin/Arzt	0 (0)	1 (2,5)
Entzug der Einverständniserklärung	3 (3,6)	3 (7,5)
Andere	2 (2,4)	0 (0)
Studienabbruch in Teil 1, n (%)	7 (8,4)	5 (12,5)
Aufgrund von:		
Nachfolgende TGCT-Therapie und/oder Operation	3 (3,6)	0 (0)
Entscheidung Ärztin/Arzt	1 (1,2)	1 (2,5)
Entzug der Einverständniserklärung	3 (3,6)	3 (7,5)
Andere	0 (0)	1 (2,5)
Übergang in Studienteil 2, n (%) <sup>3)</sup>	74 (89,2)	35 (87,5)
Mediane Behandlungsdauer, Monate (min; max)	5,5 (1; 6)	N = 39 5,5 (0; 7)
Anzahl der Behandlungszyklen, Median (min; max)	6,0 (2; 7)	N = 39 6,0 (1; 8)
Dosismodifikationen, n (%)	49 (59,0)	N = 39 5 (12,8)
Dosisreduktionen	25 (30,1)	0 (0)
Behandlungsunterbrechung	43 (51,8)	5 (12,8)
<i>Behandlungsdauer (Monate)</i>		N = 39
< 1	1 (1,2)	1 (2,6)
1 bis < 3	3 (3,6)	1 (2,6)
3 bis < 6	78 (94,0)	36 (92,3)
≥ 6	1 (1,2)	1 (2,6)
Mediane Beobachtungsdauer, Wochen (min; max) <sup>4)</sup>	24,3 (18; 47)	24,1 (0; 50)

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

<sup>2)</sup> Eine Person wurde verfrüht randomisiert und später als ungeeignet für einen Studieneinschluss erachtet.

<sup>3)</sup> Beinhaltet einen Wechsel auf Vimseltinib im Placebo-Arm.

<sup>4)</sup> Angabe aus Modul 4.

Abkürzungen: ITT: Intention to Treat; TGCT: Tenosynovialer Riesenzelltumor; UE: Unerwünschtes Ereignis.

Tabelle 10: Charakterisierung der Studienpopulation; Studie MOTION

<b>Studie MOTION</b> <b>Charakterisierung der Studienpopulation</b>	<b>Vimseltinib</b> <b>N = 83</b>	<b>Placebo</b> <b>N = 40</b>
<i>Alter (Jahre)</i> MW (SD) Median (min; max)	43,8 (13,9) 45,0 (20; 78)	42,5 (13,7) 43,0 (21; 72)
<i>Altersgruppe (Jahre), n (%)</i> ≥ 18 bis < 50 ≥ 50 bis < 65 ≥ 65 bis < 75 ≥ 75 bis < 85	54 (65,1) 24 (28,9) 3 (3,6) 2 (2,4)	27 (67,5) 9 (22,5) 4 (10,0) 0 (0)
<i>Geschlecht, n (%)</i> männlich weiblich	37 (44,6) 46 (55,4)	13 (32,5) 27 (67,5)
<i>Abstammung (genetisch), n (%)</i> asiatisch schwarz oder afroamerikanisch kaukasisch/weiß nicht berichtet unbekannt	1 (1,2) 4 (4,8) 59 (71,1) 18 (21,7) 1 (1,2)	4 (10,0) 0 (0) 21 (52,5) 13 (32,5) 2 (5,0)
<i>Region, n (%)</i> Nordamerika Asien Europa Ozeanien	14 (16,9) 0 (0) 57 (68,7) 12 (14,5)	5 (12,5) 1 (2,5) 32 (80,0) 2 (5,0)
<i>Stratifikationsfaktor „Region“, n (%)<sup>1)</sup></i> USA Rest der Welt	9 (10,8) 74 (89,2)	4 (10,0) 36 (90,0)
<i>Zeit von der ursprünglichen Diagnose bis zum Datum der ersten Dosis der Studienmedikation (Jahre)<sup>2)</sup></i> MW (SD) Median (min; max)	N = 83 5,50 (5,17) 4,31 (0,1; 22,7)	N = 39 5,63 (5,01) 4,80 (0,1; 16,5)
<i>Lage des Tumors, n (%)<sup>1)</sup></i> Untere Gliedmaßen Alle anderen Gliedmaßen	73 (88,0) 10 (12,0)	36 (90,0) 4 (10,0)
<i>Subtyp der Erkrankung, n (%)</i> Diffuser TGCT Nodulärer TGCT Keine Angabe	57 (68,7) 26 (31,3) 0 (0)	28 (70,0) 10 (25,0) 2 (5,0)
<i>Primär betroffenes Gelenk, n (%)</i> Knie Knöchel Hüfte Fuß Handgelenk Hand Schulter Ellenbogen Andere	56 (67,5) 9 (10,8) 11 (13,3) 1 (1,2) 2 (2,4) 2 (2,4) 1 (1,2) 1 (1,2) 0 (0)	27 (67,5) 6 (15,0) 1 (2,5) 3 (7,5) 1 (2,5) 0 (0) 1 (2,5) 0 (0) 1 (2,5)



<b>Studie MOTION</b> <b>Charakterisierung der Studienpopulation</b>	<b>Vimseltinib</b> <b>N = 83</b>	<b>Placebo</b> <b>N = 40</b>
<i>Anzahl der Zielläsionen zu Baseline gemäß IRR, n (%)</i> 1	83 (100)	40 (100)
<i>Stärkste Schmerzen (BPI-SF)<sup>3)</sup></i> MW (SD) Median (min; max)	5,52 (2,11) 5,71 (0,0, 9,9)	6,05 (1,78) 6,11 (1,7; 8,7)
<i>NRS-Steifigkeit<sup>3)</sup></i> MW (SD) Median (min; max)	5,1 (2,00) 5,3 (0; 8)	5,2 (1,81) 5,6 (1; 8)
<i>BMI (kg/m<sup>2</sup>), n (%)</i> MW (SD) Median (min; max)	N = 81 27,5 (5,92) 26,2 (18,8; 46,0)	27,1 (6,15) 25,9 (17,2; 44,6)
<i>Vorausgehende Behandlung des TGCT, n (%)<sup>4)</sup></i> TGCT-Operation Strahlentherapie Systemische Therapie	64 (77,1) 9 (10,8) 19 (22,9)	27 (67,5) 2 (5,0) 9 (22,5)
<i>Vorherige systemische TGCT-Therapie, n (%)</i> Jegliche vorherige systemische Therapie Imatinib Nilotinib Andere	19 (22,9) 16 (19,3) 2 (2,4) 1 (1,2)	9 (22,5) 7 (17,5) 4 (10,0) 0 (0)

<sup>1)</sup> Stratifikationsfaktor der Studie MOTION.

<sup>2)</sup> Angabe aus Modul 4 der Nutzenbewertung.

<sup>3)</sup> Einschlusskriterium für die Studie MOTION: Mind. mäßige Schmerzen oder mind. mäßige Steifigkeit (definiert als ein Wert von 4 oder mehr).

<sup>4)</sup> Erhalt mehrerer TGCT-gerichteter, vorausgehender Behandlungen möglich.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; IRR: Unabhängige radiologische Überprüfung; MW: Mittelwert; NRS: Numerische Bewertungsskala; SD: Standardabweichung; TGCT: Tenosynovialer Riesenzelltumor.

### Protokollverletzungen

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

## Begleitmedikation

Obwohl in der Studie MOTION im Vimseltinib-Arm etwas häufiger Paracetamol und Ibuprofen eingesetzt wurden, war die Begleitmedikation auf Ebene der therapeutischen Klassen ausgeglichen. Ausgenommen davon sind „Antihistaminika zur systemischen Anwendung“ und „Dermatologische Kortikosteroide“.

*Tabelle 11: Begleitmedikation bei mindestens 15 % in einem Studienarm; Studie MOTION, Teil 1*

<b>Studie MOTION</b> <b>Begleitmedikation bei mindestens 15 % in einem Studienarm</b> <i>Therapeutic Class Level 2<sup>1)</sup></i> <i>Preferred Name</i>	<b>Vimseltinib</b> <b>N = 83</b> <i>n (%)</i>	<b>Placebo</b> <b>N = 39</b> <i>n (%)</i>
Jegliche Begleitmedikation	76 (91,6)	37 (94,9)
Analgetika	51 (61,4)	23 (59,0)
Paracetamol	39 (47,0)	14 (35,9)
Antiphlogistika und Antirheumatika	43 (51,8)	18 (46,2)
Ibuprofen	27 (32,5)	8 (20,5)
Antihistaminika zur systemischen Anwendung	26 (31,3)	5 (12,8)
Dermatologische Kortikosteroide	20 (24,1)	1 (2,6)
Vitamine	17 (20,5)	6 (15,4)
Renin-Angiotensin-System-Hemmer	16 (19,3)	9 (23,1)
Sexualhormone und Modulatoren des Genitalsystems	16 (19,3)	5 (12,8)
Diuretika	15 (18,1)	2 (5,1)
Antibiotika zur systemischen Anwendung	12 (14,5)	7 (17,9)
Psycholeptika	11 (13,3)	8 (20,5)

<sup>1)</sup> Kodiert nach WHO Drug Dictionary Anatomical Therapeutic Chemical 2nd level (ATC-2) and preferred name (Stand: März 2023 Global B3 oder später).

Abkürzungen: WHO: World Health Organization.

## 3.2 Mortalität

In der randomisierten Phase der Studie MOTION sind keine Todesfälle aufgetreten.

### 3.3 Morbidität

#### Körperliche Funktionsfähigkeit mittels PROMIS PF

Für die primär betrachtete Analyse des Effekts zu Woche 25 liegen keine Tests auf Interaktion vor. Die Effektschätzer in den Subgruppen für die Effekte zu Woche 25 lassen keine relevanten Unterschiede zwischen den betrachteten Gruppen erkennen.

Zudem liegen für den PROMIS-PF a priori geplante Sensitivitätsanalysen im Hinblick auf fehlende Werte vor, die jedoch mit den Ergebnissen aus Tabelle 15 vergleichbar sind.

*Tabelle 12: PROMIS PF, MMRM-Analysen anhand beobachteter Werte zu Woche 25;  
Studie MOTION, ITT-Population*

<b>Studie MOTION PROMIS PF<sup>1)</sup></b>	<b>Vimseltinib N = 83</b>	<b>Placebo N = 40</b>
<i>Baseline</i>		
Verfügbare Werte, n (%)	83 (100)	40 (100)
MW (SD)	39,0 (6,14)	38,5 (5,98)
<i>Woche 25</i>		
Verfügbare Werte, n (%)	63 (75,9)	30 (75,0)
MW (SD)	43,7 (6,13)	40,7 (6,74)
<i>Veränderung zu Woche 25</i>		
Verfügbare Werte, n (%)	63 (75,9)	30 (75,0)
LS Mean (SE) <sup>2)</sup>	4,6 (0,96)	1,3 (0,88)
<i>Gesamteffekt der Behandlung bis Woche 25</i>		
Verfügbare Werte, n (%)	83 (100)	37 (92,5)
LS Mean (SE) <sup>2)</sup>	3,9 (0,84)	1,2 (0,84)
<i>Veränderung im Gesamteffekt</i>		
LS-Mean-Differenz [95%-KI] <sup>2)</sup> ; p-Wert	2,7 [1,2; 4,2]; 0,0005	
Hedges' g [95%-KI]	0,701 [0,306; 1,096]	
<i>Veränderung zu Woche 25</i>		
LS-Mean-Differenz [95%-KI] <sup>2)</sup> ; p-Wert	3,3 [1,4; 5,2]; 0,0007	
Hedges' g [95%-KI]	k. A. <sup>3)</sup>	

<sup>1)</sup> T-transformierte Werte, MW = 50, SD = 10. Höhere Werte bedeuten eine bessere körperliche Funktionsfähigkeit.

<sup>2)</sup> MMRM unter Verwendung des Sandwich-Schätzers zur Schätzung der Varianz-Kovarianz-Matrix. Die abhängige Variable war dabei die Veränderung gegenüber dem Ausgangswert. Jedes dieser Modelle enthielt feste Effekte für die Behandlungsgruppe, den Zeitpunkt sowie die Interaktion zwischen Behandlungsgruppe und Zeitpunkt. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorlokalisation“ und „Region“. Zudem wurde für den Ausgangswert des entsprechenden Endpunkts als Co-Faktor adjustiert.

<sup>3)</sup> Auf Basis der Ergebnisse des Hedges' g für den Gesamteffekt bis Woche 25, erscheint das Vorliegen eines klinisch relevanten Effekts auch für die Differenz der Veränderung zu Woche 25 plausibel.

Abkürzungen: ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MMRM: Mixed Model for Repeated Measures; MW: Mittelwert; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; SD: Standardabweichung; SE: Standardfehler.

### Steifigkeit mittels NRS

Für den Endpunkt „Steifigkeit mittels NRS“ liegen keine relevanten, statistisch signifikanten Interaktionsterme vor.

*Tabelle 13: Anteil der Personen mit einer Verbesserung um mindestens 2 Punkte der NRS-Steifigkeit zu Woche 25; Studie MOTION, ITT-Population*

Studie MOTION NRS-Steifigkeit Responderanalyse zu Woche 25 <sup>1)</sup>	Vimseltinib N = 83	Placebo N = 40
<b>Verbesserung um mindestens 2 Punkte</b>		
Verfügbare Werte, n (%) <sup>2)</sup>	63 (75,9)	27 (67,5)
Responder, n (%)	32 (38,6)	6 (15,0)
RR [95%-KI]; p-Wert <sup>3)</sup>	2,65 [1,21; 5,82]; 0,01	

<sup>1)</sup> Die Erhebungen für den angegebenen Zeitpunkt „zu Woche 25“ erfolgten in Woche 2 und 3 des letzten Behandlungszyklus.

<sup>2)</sup> Die Vollständigkeit der Erhebungen zur NRS-Steifigkeit ist unklar. Regelmäßig erfolgte eine Erhebung an jedem zweiten Tag, wobei jeweils nur Woche 2 und 3 des jeweiligen Behandlungszyklus über einen gleichgewichteten Durchschnitt gewertet wurden. Es mussten mind. 4 von 7 möglichen Messzeitpunkten vorhanden sein.

<sup>3)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Werts wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorklassifikation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

Abkürzungen: ITT: Intention to Treat; KI: Konfidenzintervall; NRS: Numerische Bewertungsskala; RR: Relatives Risiko.

### Schmerzen mittels BPI-SF

Für den BPI-SF liegen keine stetigen Auswertungen für die Veränderung zu Woche 25 vor. Es sind lediglich Post-hoc-Auswertungen für den Gesamteffekt der Behandlung verfügbar, die für beide Items einen statistisch signifikanten Unterschied zwischen den Behandlungsarmen zeigen. Auf Basis der Ergebnisse des Hedges'g erscheint das Vorliegen eines klinisch relevanten Effekts für den Gesamteffekt zu Woche 25 plausibel.

Für den BPI-SF liegen statistisch signifikante Interaktionsterme für „Lage des Tumors“, „Erkrankung an den großen vs. kleinen Gelenken“ sowie „Erkrankung im Bereich des Knies“ vor. In allen Subgruppen zeigte sich jedoch ein statistisch signifikanter Vorteil von Vimseltinib, der gemessen am Hedges'g auch auf einen klinisch relevanten Effekt hindeutet.

*Tabelle 14: Schmerzen mittels BPI-SF, MMRM-Analysen anhand beobachteter Werte zu Woche 25; Studie MOTION, ITT-Population*

Studie MOTION Schmerzen mittels BPI-SF <sup>1)</sup>	Vimseltinib N = 83	Placebo N = 40
<b>Stärkster Schmerz</b>		
<i>Baseline</i>		
Verfügbare Werte, n (%) <sup>2)</sup>	83 (100)	40 (100)
MW (SD)	5,5 (2,11)	6,0 (1,78)
<i>Woche 25</i>		
Verfügbare Werte, n (%) <sup>2)</sup>	68 (81,9)	31 (77,5)
MW (SD)	3,3 (2,23)	4,9 (1,93)

Studie MOTION Schmerzen mittels BPI-SF <sup>1)</sup>	Vimseltinib N = 83	Placebo N = 40
Veränderung zu Woche 25 Verfügbare Werte, n (%) <sup>2)</sup> LS Mean (SE) <sup>3)</sup>	68 (81,9) -2,6 (0,33)	31 (77,5) -1,0 (0,39)
Gesamteffekt der Behandlung bis Woche 25 Verfügbare Werte, n (%) <sup>2)</sup> LS Mean (SE) <sup>3)</sup>	83 (100) -2,0 (0,30)	39 (97,5) -0,8 (0,34)
Unterschied im Gesamteffekt der Behandlung LS-Mean-Differenz [95%-KI] <sup>3)</sup> ; p-Wert Hedges' g [95%-KI]	-1,2 [-1,8; -0,6]; < 0,0001 -0,79 [-1,18; -0,40]	
Veränderung zu Woche 25 LS-Mean-Differenz [95%-KI]; p-Wert Hedges' g [95%-KI]	k. A.	
Durchschnittlicher Schmerz		
Baseline Verfügbare Werte, n (%) <sup>2)</sup> MW (SD)	83 (100) 4,62 (2,00)	40 (100) 5,19 (1,96)
Woche 25 Verfügbare Werte, n (%) <sup>2)</sup> MW (SD)	68 (81,9) 2,54 (2,07)	31 (77,5) 4,09 (2,05)
Veränderung zu Woche 25 Verfügbare Werte, n (%) <sup>2)</sup> LS Mean (SE) <sup>3)</sup>	68 (81,9) -2,4 (0,30)	31 (77,5) -1,0 (0,36)
Gesamteffekt der Behandlung bis Woche 25 Verfügbare Werte, n (%) <sup>2)</sup> LS Mean (SE) <sup>3)</sup>	83 (100) -1,9 (0,27)	39 (97,5) -0,9 (0,30)
Unterschied Gesamteffekt der Behandlung LS-Mean-Differenz [95%-KI] <sup>3)</sup> ; p-Wert Hedges' g [95%-KI]	-1,0 [-1,6; -0,5]; 0,0001 -0,76 [-1,15; -0,37]	
Veränderung zu Woche 25 LS-Mean-Differenz [95%-KI]; p-Wert Hedges' g [95%-KI]	k. A.	

<sup>1)</sup> Es wurden ausschließlich die Items „Stärkster Schmerz“ und „Durchschnittlicher Schmerz“ erhoben. Skala von 0 bis 10, wobei niedrigere Werte weniger Schmerzen bedeuten. Die Erhebungen für den angegebenen Zeitpunkt „zu Woche 25“ erfolgten in Woche 2 und 3 des letzten Behandlungszyklus.

<sup>2)</sup> Die Erhebung erfolgte täglich. Die Vollständigkeit der Tagebucheinträge ist unklar. Zur Auswertung wurden die Werte aller verfügbaren Erhebungen in Woche 2 und 3 eines jeden Zyklus gemittelt; es mussten mind. 4 von 14 Werten verfügbar sein, damit eine Person als nicht-fehlend gilt.

<sup>3)</sup> MMRM unter Verwendung des Sandwich-Schätzers zur Schätzung der Varianz-Kovarianz-Matrix. Die abhängige Variable war dabei die Veränderung gegenüber dem Ausgangswert. Jedes dieser Modelle enthielt feste Effekte für die Behandlungsgruppe, den Zeitpunkt sowie die Interaktion zwischen Behandlungsgruppe und Zeitpunkt. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorlokalisation“ und „Region“. Zudem wurde für den Ausgangswert des entsprechenden Endpunkts als Co-Faktor adjustiert.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; LS: Least Squares; MMRM: Mixed Model for Repeated Measures; MW: Mittelwert; SD: Standardabweichung; SE: Standardfehler.

### Allgemeiner Gesundheitszustand mittels EQ-5D-VAS

Für den Endpunkt „EQ-5D-VAS“ zeigten sich keine statistisch signifikanten Interaktionsterme.

*Tabelle 15: Anteil der Personen mit einer Verbesserung um mindestens 15 % der Skalenspannweite der EQ-5D-VAS zu Woche 25; Studie MOTION, ITT-Population*

Studie MOTION EQ-5D-VAS Responderanalyse zu Woche 25	Vimseltinib N = 83	Placebo N = 40
<b>Verbesserung um mindestens 15 % der Skalenspannweite</b>		
Verfügbare Werte, n (%)	64 (77,1)	30 (75,0)
Responder, n (%)	26 (31,3)	4 (10,0)
RR [95%-KI]; p-Wert <sup>1)</sup>	3,20 [1,21; 8,44]; 0,02	

<sup>1)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Werts wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorklassifikation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

Abkürzungen: EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; ITT: Intention to Treat; KI: Konfidenzintervall; RR: Relatives Risiko.

### Körperliche Funktionsfähigkeit / Bewegungsumfang mittels PGIS (ergänzend dargestellt)

*Tabelle 16: Anteil der Personen mit einer Verbesserung im PGIS zu Woche 25 (ergänzend dargestellt); Studie MOTION, ITT-Population*

Studie MOTION PGIS (ergänzend dargestellt) Responderanalyse zu Woche 25	Vimseltinib N = 83	Placebo N = 40
Verbesserung im PGIS-ROM <sup>1)</sup>		
Verfügbare Werte, n (%)	63 (75,9)	30 (75,0)
Responder, n (%)	32 (38,6)	8 (20,0)
RR [95%-KI]; p-Wert <sup>2)</sup>	1,93 [0,98; 3,79]; 0,06	
Verbesserung im PGIS-PF <sup>1)</sup>		
Verfügbare Werte, n (%)	63 (75,9)	30 (75,0)
Responder, n (%)	36 (43,4)	8 (20,0)
RR [95%-KI]; p-Wert <sup>2)</sup>	2,10 [1,09; 4,06]; 0,03	

<sup>1)</sup> Verbesserung war definiert als Verbesserung um mind. 1 Punkt.

<sup>2)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Werts wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorklassifikation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

Abkürzungen: ITT: Intention to Treat; KI: Konfidenzintervall; PF: Körperliche Funktionsfähigkeit; PGIS: Patient Global Impression of Severity; ROM: Bewegungsumfang; RR: Relatives Risiko.

### Allgemeine Krankheitssymptomatik mittels PGIC

Für den Endpunkt „PGIC-Krankheitssymptomatik“ zeigten sich keine statistisch signifikanten Interaktionsterme.

*Tabelle 17: Anteil der Personen mit einer Verbesserung im PGIC-Krankheitssymptomatik zu Woche 25; Studie MOTION, ITT-Population*

Studie MOTION PGIC-Krankheitssymptomatik Responderanalyse zu Woche 25	Vimseltinib N = 83	Placebo N = 40
<b>Verbesserung im PGIC-Krankheitssymptomatik<sup>1)</sup></b>		
Verfügbare Werte, n (%)	64 (77,1)	30 (75,0)
Responder, n (%)	55 (66,3)	9 (22,5)
RR [95%-KI]; p-Wert <sup>2)</sup>	2,95 [1,63; 5,35]; 0,0004	

<sup>1)</sup> Verbesserung war definiert als „sehr viel besser“, „viel besser“ und „wenig besser“.

<sup>2)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Werts wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorlokalisation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

Abkürzungen: ITT: Intention to Treat; KI: Konfidenzintervall; PGIC: Patient Global Impression of Change; RR: Relatives Risiko.

## 3.4 Lebensqualität

Es liegen keine Daten zur Lebensqualität vor.

## 3.5 Sicherheit

Die mediane Behandlungsdauer ist in beiden Studienarmen mit 5,5 Monaten gleich. Der pU legte keine zusätzlichen Auswertungen unter Nichtberücksichtigung erkrankungsbezogener Ereignisse bzw. Ereignissen der Grunderkrankung vor. Es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung mit in die Erfassung der UE eingingen. Allerdings durften Symptome der untersuchten Krankheit nicht als UE erfasst werden, solange sie innerhalb der normalen täglichen Schwankungen oder des erwarteten Krankheitsverlaufs liegen.

In beiden Behandlungsgruppen gingen knapp 90 % der Personen in Studienteil 2 über, sodass die 30-tägige Sicherheitsnachbeobachtung offensichtlich nur für einen kleinen Teil der Studienpopulation in die vorliegenden Sicherheitsauswertungen einging.

**Tabelle 18: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase, Studie MOTION, Sicherheitspopulation**

<b>MOTION</b> <b>Zusammenfassung der UE</b> <i>Personen mit mindestens einem ...</i>	<b>Vimseltinib</b> <b>N = 83</b> <i>n (%)</i>	<b>Placebo</b> <b>N = 39</b> <i>n (%)</i>	<b>RR [95%-KI]</b> <b>p-Wert<sup>1)</sup></b>
UE (ergänzend dargestellt)	83 (100)	37 (94,9)	-
schweren UE (CTCAE-Grad $\geq 3$ )	31 (37,3)	4 (10,3)	3,64 [1,38; 9,60] 0,01
SUE	6 (7,2)	1 (2,6)	2,82 [0,35; 22,62] 0,33
UE, das zum Abbruch der Studienmedikation führte	5 (6,0)	0 (0)	5,24 [0,30; 92,4] 0,26

<sup>1)</sup> Log-Binomial-Regression mit dem Term „Behandlung“.

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

Für die schweren UE zeigten die durchgeführten Subgruppenanalysen in Modul 4 eine signifikante Interaktion für eine vorherige Imatinib- oder Nilotinib-Behandlung ( $p = 0,009$ ). Bei Personen, die nicht mit Nilotinib oder Imatinib vorbehandelt waren, ergab sich ein statistisch signifikanter Nachteil für Vimseltinib (RR: 12,2 (95%-KI: [1,74; 85,6]);  $p = 0,01$ ). Bei Personen mit Nilotinib- oder Imatinib-Vorbehandlung war der Effekt nicht statistisch signifikant (RR: 0,79 (95%-KI: [0,24; 2,60]),  $p = 0,70$ ).

### **Unerwünschte Ereignisse**

**Tabelle 19: UE mit Inzidenz  $\geq 10$  % während der verblindeten Behandlungsphase; Studie MOTION, Sicherheitspopulation**

<b>Studie MOTION</b> <b>UE</b> <b>MedDRA-Systemorganklasse</b> <i>Preferred Term</i>	<b>Vimseltinib</b> <b>N = 83</b> <i>n (%)</i>	<b>Placebo</b> <b>N = 39</b> <i>n (%)</i>	<b>RR [95%-KI]</b> <b>p-Wert<sup>1)</sup></b>
<b>Augenerkrankungen</b>	61 (73,5)	9 (23,1)	3,19 [1,77; 5,73] 0,0001
Periorbitalödem	37 (44,6)	5 (12,8)	3,48 [1,48; 8,16] 0,004
Augenlidödem	11 (13,3)	2 (5,1)	2,58 [0,60; 11,1] 0,20
Tränensekretion verstärkt	10 (12,0)	0 (0)	10,0 [0,60; 166,4] 0,11
<b>Erkrankungen des Gastrointestinaltrakts</b>	32 (38,6)	13 (33,3)	1,16; [0,69; 1,95] 0,58
Übelkeit	21 (25,3)	8 (20,5)	1,23 [0,60; 2,53] 0,57
Diarrhö	10 (12,0)	8 (20,5)	0,59; [0,25; 1,37] 0,22



Studie MOTION UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Vimseltinib N = 83 n (%)	Placebo N = 39 n (%)	RR [95%-KI] p-Wert <sup>1)</sup>
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	64 (77,1)	18 (46,2)	1,67 [1,17; 2,39] 0,005
Ermüdung	27 (32,5)	6 (15,4)	2,11 [0,95; 4,70] 0,07
Asthenie	22 (26,5)	9 (23,1)	1,15 [0,58; 2,26] 0,69
Gesichtsödem	26 (31,3)	3 (7,7)	4,07 [1,31; 12,6] 0,02
Ödem peripher	15 (18,1)	3 (7,7)	2,35 [0,72; 7,64] 0,16
Generalisiertes Ödem	11 (13,3)	0 (0)	11,0 [0,66; 181,2] 0,09
<b>Infektionen und parasitäre Erkrankungen</b>	29 (34,9)	8 (20,5)	1,70 [0,86; 3,38] 0,13
COVID-19	10 (12,0)	1 (2,6)	4,70 [0,62; 35,4] 0,13
<b>Untersuchungen</b>	40 (48,2)	4 (10,3)	4,70 [1,81; 12,2] 0,002
Aspartat-Aminotransferase erhöht	19 (22,9)	1 (2,6)	8,93 [1,24; 64,3] 0,03
Kreatinphosphokinase im Blut erhöht	20 (24,1)	0 (0)	19,5 [1,21; 314,7] 0,04
Alanin-Aminotransferase erhöht	9 (10,8)	1 (2,6)	4,23 [0,56; 32,2] 0,16
<b>Stoffwechsel- und Ernährungsstörungen</b>	18 (21,7)	3 (7,7)	2,82 [0,88; 9,01] 0,08
<b>Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen</b>	29 (34,9)	14 (35,9)	0,97 [0,58; 1,62] 0,92
Arthralgie	16 (19,3)	6 (15,4)	1,25 [0,53; 2,95] 0,61
<b>Erkrankungen des Nervensystems</b>	37 (44,6)	13 (33,3)	1,3 [0,81; 2,22] 0,26
Kopfschmerzen	23 (27,7)	10 (25,6)	1,08; [0,57; 2,04] 0,81
<b>Psychiatrische Erkrankungen</b>	7 (8,4)	3 (7,7)	1,10 [0,30; 4,01] 0,89
<b>Erkrankungen der Atemwege, des Brustraums und Mediastinums</b>	12 (14,5)	2 (5,1)	2,82 [0,66; 12,0] 0,16
<b>Erkrankungen der Haut und des Unterhautgewebes</b>	48 (57,8)	6 (15,4)	3,76 [1,76; 8,03] 0,0006
Pruritus	24 (28,9)	3 (7,7)	3,76 [1,20; 11,7] 0,02

Studie MOTION UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Vimseltinib N = 83 n (%)	Placebo N = 39 n (%)	RR [95%-KI] p-Wert <sup>1)</sup>
Ausschlag	16 (19,3)	2 (5,1)	3,76 [0,91; 15,6] 0,07
Ausschlag makulo-papulös	16 (19,3)	0 (0)	15,71 [0,97; 255,4] 0,05
<b>Gefäßkrankungen</b>	17 (20,5)	7 (17,9)	1,14 [0,52; 2,52] 0,74
Hypertonie	14 (16,9)	4 (10,3)	1,65 [0,58; 4,67] 0,35

<sup>1)</sup> Log-Binomial-Regression mit dem Term „Behandlung“.

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; KI: Konfidenzintervall; RR: Relatives Risiko; UE: Unerwünschtes Ereignis.

### Schwere unerwünschte Ereignisse

*Tabelle 20: Schwere UE (CTCAE-Grad  $\geq 3$ ), die bei  $\geq 5$  % der Personen in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term, während der verblindeten Behandlungsphase; Studie MOTION, Sicherheitspopulation*

Studie MOTION Schwere UE <i>MedDRA-Systemorganklasse</i> <i>Preferred Term</i>	Vimseltinib N = 83 n (%)	Placebo N = 39 n (%)	RR [95%-KI] p-Wert <sup>1)</sup>
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	4 (4,8)	2 (5,1)	0,94 [0,18; 4,91] 0,94
<b>Infektionen und parasitäre Erkrankungen</b>	5 (6,0)	0 (0)	5,24 [0,30; 92,4] 0,26
<b>Untersuchungen</b>	9 (10,8)	0 (0)	9,05 [0,54; 151,6] 0,13
Kreatinphosphokinase im Blut erhöht	8 (9,6)	0 (0)	8,10 [0,48; 136,8] 0,15

<sup>1)</sup> Log-Binomial-Regression mit dem Term „Behandlung“.

Abkürzungen: MedDRA: Medical Dictionary for Regulatory Activities; KI: Konfidenzintervall; RR: Relatives Risiko; UE: Unerwünschtes Ereignis.

### Schwerwiegende unerwünschte Ereignisse / Unerwünschte Ereignisse von besonderem Interesse

Im Rahmen der randomisierten Phase der Studie MOTION traten keine SUE auf Ebene der SOC und PT auf, welche bei mindestens 5 % der Patientinnen und Patienten in einem Studienarm aufgetreten sind. Zudem wurden in der Studie MOTION keine UE von besonderem Interesse bestimmt.

## 4 Diskussion der Methodik und Ergebnisse

### 4.1 Design und Methodik der Studie

#### Studiendesign

Bei der Studie MOTION handelt es sich um eine randomisierte, multizentrische, Placebo-kontrollierte Studie zur Untersuchung der Wirksamkeit und Sicherheit von Vimseltinib bei Personen mit tenosynovialen Riesenzelltumoren (TGCT). Primärer Endpunkt war die „Objektive Ansprechrare“. Die Studie besteht aus einem 24-wöchigem randomisierten, doppelblinden Teil 1 und einem offenen, einarmigen Teil 2, wobei Personen aus dem Placebo-Arm in Teil 2 auf Vimseltinib wechseln konnten. Die Randomisierung erfolgte im Verhältnis 2:1 auf die Behandlungsarme Vimseltinib oder Placebo, stratifiziert nach „Tumorklassifikation“ (untere Gliedmaßen vs. alle anderen) und „Region“ (USA vs. Rest der Welt). An Teil 1 schließt sich ein einarmiger Teil 2 der Behandlungsphase bis Woche 49 an sowie eine Erweiterungsphase bis Krankheitsprogression oder Studienende. Das Studienende der noch laufenden Studie ist für Juli 2026 geplant.

In der vorliegenden Nutzenbewertung werden ausschließlich die Ergebnisse des bereits abgeschlossenen, randomisierten Studienteils 1 berücksichtigt, da die Dauer mit 24 Wochen als ausreichend erachtet wird und die darüber hinausgehenden Ergebnisse nicht vergleichend sind. Die Ergebnisse werden im Rahmen eines Datenschnitts vom 22.08.2023 berichtet, der dem Abschluss des vergleichenden Studienteil 1 entspricht.

Nach Einschluss der ersten Person in die Studie wurden 2 Amendments am Studienprotokoll vorgenommen. Obwohl sich die Änderungen auch auf die Ein- und Ausschlusskriterien, die zulässigen bzw. nicht-erlaubten Begleitmedikationen sowie den Umgang mit Dosisunterbrechungen beziehen, erscheinen die sich daraus für die Studie ergebenden Auswirkungen gering.

#### Studienpopulation

In die Studie konnten Personen ab 18 Jahren mit einer histologisch bestätigten Diagnose von TGCT eingeschlossen werden, bei denen eine chirurgische Resektion zu einer Verschlechterung der Funktionseinschränkungen oder einer schweren Morbidität führen würde. Die Patientinnen und Patienten mussten eine symptomatische Erkrankung mit mindestens mäßigen Schmerzen oder mäßiger Steifigkeit aufweisen mit messbarer Erkrankung gemäß RECIST mit mindestens einer Läsion größer als 2 cm. Eine vorherige systemische Therapie, die auf CSF1 oder CSF1R abzielt, war verboten, vorherige Therapien mit Imatinib, Nilotinib oder chirurgische Resektionen waren jedoch erlaubt.

Von den 146 gescreenten Personen konnten 123 in die Studie aufgenommen werden, wobei 83 in die Interventionsgruppe randomisiert wurden und 40 in die Kontrollgruppe. Davon wurde eine Person aus dem Placebo-Arm nicht behandelt, da sie für einen Studieneinschluss nicht geeignet war und verfrüht randomisiert wurde. Studienabbrüche in Teil 1 der Studie MOTION sind für lediglich 7 Personen (8 %) aus dem Interventionsarm und 5 (13 %) aus dem Kontrollarm dokumentiert. Nach Teil 1 gingen knapp 90 % aus beiden Studienarmen in Studienteil 2 über und wurden dort weiterbehandelt. Eine Erkrankungsprogression ist lediglich für 1 Person aus dem Placebo-Arm dokumentiert, die daraufhin die Behandlung abbrach und direkt in Studienteil 2 auf Vimseltinib wechselte.

Lediglich 10 % der Studienpopulation kamen aus den USA und 90 % aus dem Rest der Welt (gleichzeitig ein Stratifikationskriterium). 72 % der Studienpopulation wurde an Studienzentren in Europa behandelt, davon 7 % in Studienzentren in Deutschland. Das Alter der Studienpopulation lag im Median bei 44 Jahren, wobei der Anteil der Studienteilnehmerinnen etwas größer war. Im Mittel vergingen über 5,5 Jahre von der Diagnosestellung bis zur ersten Dosis mit

der Studienmedikation. Bei knapp 90 % der Studienteilnehmenden waren die unteren Gliedmaßen betroffen und hier dominierte mit 68% das Knie. Fast 70 % litten unter einem diffusen Subtyp. Zum Anteil der Personen, die bereits eine Tumor-gerichtete Therapie erhalten hatten und damit vorbehandelt waren, liegen keine Daten vor. 77 % der Personen aus der Interventionsgruppe und 68 % aus der Kontrollgruppe wurden bereits mindestens einer TGCT-Operation unterzogen, 23 % aus beiden Studienarmen erhielten bereits eine systemische Therapie und 11 % bzw. 5 % waren mit einer Strahlentherapie vorbehandelt.

Die Patientencharakteristika waren zwischen den Behandlungsarmen weitgehend balanciert. In der Kontrollgruppe war der Frauenanteil mit 68 % etwas höher als in der Interventionsgruppe mit 55 %. Zudem war die Hüfte im Vimseltinib-Arm etwas häufiger als primär betroffenes Gelenk (13 vs. 2,5 %) angegeben, dafür waren im Placebo-Arm der Knöchel (11 vs. 15 %) sowie der Fuß (1 vs. 8 %) etwas häufiger betroffen.

### **Studienmedikation**

Vimseltinib wurde als Dauertherapie zweimal wöchentlich in einer Dosis von 30 mg jeweils an Tag 1 und Tag 5 verabreicht. Die Einnahme sollte immer zur gleichen Tageszeit auf nüchternen Magen erfolgen. Die Verabreichung von Placebo war vergleichbar, nur ohne den aktiven Wirkstoff. Obwohl es sich um eine Dauertherapie handelt, wurde in der Studie MOTION in Behandlungszyklen bestehend aus 28 Tagen gerechnet. Im Studienprotokoll sind eine Vielzahl von Regelungen für Dosisanpassungen spezifiziert, z. B. bei dermatologischen Ereignissen oder Änderungen des QTcF-Intervalls. Im Fall von UE CTCAE-Grad 4 war ein Abbruch der Studienmedikation vorgesehen. Von den eingeschlossenen Personen erhielten alle ihre zugewiesene Medikation, abgesehen von einer Person aus dem Placebo-Arm, die verfrüht randomisiert und nicht behandelt wurde.

Mit 12 % aus dem Interventionsarm und 13 % aus dem Kontrollarm war der Anteil der Personen mit einem Abbruch der Studienmedikation gering. Häufigster Grund für einen Abbruch der Studienmedikation im Interventionsarm war ein UE bei 4 Personen (5 %); beim Endpunkt „Unerwünschte Ereignisse“ sind Behandlungsabbrüche aufgrund von UE jedoch für 5 Personen (6 %) dokumentiert. Die Ursache für diese Diskrepanz ist unklar. Dosismodifikationen sind dagegen für 60 % der Patientinnen und Patienten unter Vimseltinib dokumentiert, darunter Dosisreduktionen bei 30 % und Behandlungsunterbrechungen bei 52 % der Studienpopulation. Im Placebo-Arm ereigneten sich lediglich Behandlungsunterbrechungen bei 13 % und keine Dosisreduktionen.

## **4.2 Zulassungsstatus und Zulassungspopulation**

### **Zulassungsstatus und Anwendungsgebiet**

Vimseltinib ist zugelassen zur Behandlung von erwachsenen Patientinnen und Patienten mit symptomatischen tenosynovialen Riesenzelltumoren (TGCT), die mit einer klinisch relevanten Verschlechterung der körperlichen Funktionsfähigkeit assoziiert sind und bei denen chirurgische Optionen ausgeschöpft sind oder zu einer inakzeptablen Morbidität oder Behinderung führen würden. Gemäß Fachinformation beträgt die zugelassene Dosierung 30 mg zweimal wöchentlich im Abstand von mindestens 72 Stunden, wobei die Therapie so lange fortgesetzt werden kann, wie die Patientin / der Patient davon einen Nutzen hat und keine inakzeptable Toxizität auftritt [5].

Der Einsatz von Vimseltinib in der Studie MOTION erfolgte weitestgehend Fachinformation-konform. In der Studie waren die Patientinnen und Patienten angewiesen die Medikation immer zur gleichen Tageszeit auf nüchternen Magen einzunehmen. Entsprechende Anweisungen sind in der Fachinformation nicht enthalten. Die daraus resultierenden Unsicherheiten werden jedoch als geringfügig erachtet.

Die in der Studie MOTION eingeschlossenen Patientinnen und Patienten sind vom vorliegenden Anwendungsgebiet abgedeckt. Einschlusskriterien waren eine symptomatische Erkrankung mit mindestens mäßigen Schmerzen oder mäßiger Steifigkeit sowie ein Tumor, bei dem eine Resektion möglicherweise zu einer Verschlechterung der Funktionseinschränkung oder einer schweren Morbidität führen würde. 74 % der Studienteilnehmenden wurden in der Behandlungshistorie bereits einer TGCT-assoziierten Operation unterzogen. Da zudem ein Anteil der Patientinnen und Patienten als allgemein inoperabel eingestuft wird, insbesondere Personen mit diffuser TGCT, erscheint plausibel, dass eine Operation bei den Studienteilnehmenden nicht angezeigt ist.

### **Übertragbarkeit auf deutschen Versorgungskontext**

Beim vorliegenden Anwendungsgebiet handelt es sich um eine sehr heterogene Patientenpopulation. Dies betrifft insbesondere die Lage des Tumors bzw. das betroffene Gelenk und zum anderen den Subtyp der Erkrankung (diffus oder nodulär) sowie mögliche Vorbehandlungen. Primär betroffenes Gelenk in der Studie MOTION war mit 68 % überwiegend das Knie. Tumore in den oberen Gliedmaßen, wie Ellenbogen, Schulter, Hand oder Handgelenk, waren selten (jeweils zwischen 1 und 2,5 %). Zwar ist die Verteilung der betroffenen Gelenke ähnlich zu denen, die in anderen Publikationen zu TGCT beschrieben ist [18,20], dennoch bleiben aufgrund dieser geringen Fallzahlen Unsicherheiten hinsichtlich der Übertragbarkeit der Wirksamkeitsergebnisse auf Patientinnen und Patienten mit Tumoren in den oberen Gliedmaßen. Weiterhin kann dem Studienbericht entnommen werden, dass 72 % der Studienpopulation an Studienzentren in Europa behandelt wurden, jedoch lediglich 7 % in Studienzentren in Deutschland.

Insgesamt kann von der Übertragbarkeit der Ergebnisse auf den deutschen Versorgungskontext ausgegangen werden.

### **4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit**

Zusammenfassend ergibt sich für Teil 1 der Studie MOTION ein niedriges Verzerrungspotential auf Studienebene aufgrund des randomisierten, doppelblinden Studiendesigns.

Die Ergebnissicherheit wird jedoch eingeschränkt durch die heterogene Patientenpopulation im Anwendungsgebiet in Verbindung mit der relativ geringen Patientenzahl in der Studie MOTION. Gerade Patientinnen und Patienten mit Tumoren in den oberen Extremitäten sind nur in geringem Umfang in der Studie eingeschlossen. Die Dauer der randomisierten Behandlungsphase ist mit 24 Wochen ausreichend zur Berücksichtigung der Ergebnisse aus der Studie MOTION, auch wenn ein längerer vergleichender Behandlungszeitraum wünschenswert gewesen wäre. Die Ergebnisse von Teil 2 der Studie MOTION wurden aufgrund des unkontrollierten Studiendesigns nicht berücksichtigt.

Auf Endpunktebene wird das Verzerrungspotential für die Morbiditätsendpunkte und Teile der Sicherheit jedoch als hoch eingeschätzt, aufgrund des hohen Anteils fehlender Werte und einer möglichen Entblindung eines signifikanten Teils der Interventionsgruppe. Die Ergebnissicherheit wird durch das hohe Verzerrungspotential auf Endpunktebene eingeschränkt, allerdings sind die beobachteten Effekte bei den Endpunkten der „Morbidität“ und „Sicherheit“ sehr groß. Endpunkte der Kategorie „Lebensqualität“ wurden in der Studie MOTION nicht erhoben.

Von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist auszugehen.

## 4.4 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserfassung dokumentiert. Im Verlauf der randomisierten Phase der Studie MOTION ist keine Person verstorben. Eine abschließende Bewertung der Mortalität ist aufgrund der relativ kurzen randomisierten Behandlungsphase nicht möglich.

## 4.5 Morbidität

Für die Endpunktkategorie „Morbidität“ legte der pU in Modul 4 Daten zu den Endpunkten „Tumoransprechen“ und „Bewegungsumfang“ sowie zu den patientenberichteten Endpunkten anhand der Instrumente „PROMIS PF“, „NRS-Steifigkeit“, „BPI-SF“, „EQ-5D-VAS“, „PGIS“ und „PGIC“ vor. Die detaillierte Bewertung der Endpunkte ist Kapitel 2.3 zu entnehmen. Zusammenfassend wurden für die Nutzenbewertung die Ergebnisse der patientenberichteten Endpunkte anhand der Instrumente „PROMIS PF“, „NRS-Steifigkeit NRS“, „BPI-SF“, „EQ-5D-VAS“ und „PGIC-Krankheitssymptomatik“ herangezogen. Der Endpunkt „Bewegungsumfang“ wird aufgrund unklarer Patientenrelevanz ergänzend dargestellt. Eine ergänzende Darstellung erfolgt auch für den PGIS, hier jedoch aufgrund unklarer Validität und Doppelerfassung (siehe Kapitel 2.3). Dies ist auch der Fall für Item 1 (ROM) und Item 2 (PF) des PGIC, die aufgrund der fehlenden Baseline-Erhebung und der Doppelerfassung zum PGIS im Anhang aus Transparenzgründen dargestellt werden.

Für den Endpunkt körperliche Funktionsfähigkeit anhand des PROMIS PF wurden ausschließlich die stetigen Auswertungen berücksichtigt, da das vom pU post hoc gewählte Responsekriterium von 8,025 Punkten als nicht adäquat eingestuft wurde. Bei der Veränderung zu Woche 25 im Vergleich zu Baseline zeigte sich im Interventionsarm eine deutlich stärkere Verbesserung von 4,6 Punkten im Vergleich zu Placebo mit 1,3 Punkten. Im MMRM war dieser Unterschied statistisch signifikant (LS-Mean-Differenz: 3,3 (95%-KI: [1,4; 5,2])). Für die Veränderung zu Woche 25 liegt kein Hedges' g (einem möglichen Hilfskriterium zur Einschätzung klinischer Relevanz) vor, aber auf Basis der Ergebnisse des Hedges' g für den Gesamteffekt bis Woche 25 erscheint das Vorliegen eines klinisch relevanten Effekts auch für die Differenz der Veränderung zu Woche 25 plausibel.

Hinsichtlich der Auswertung der NRS für die stärkste Steifigkeit des betroffenen Gelenks in den letzten 24 Stunden wurde für die Nutzenbewertung die Responderanalyse mit einer a priori definierten Responseschwelle von Verbesserungen um 2 Punkten berücksichtigt. 39 % der Patientinnen und Patienten des Vimseltinib-Arms erfüllten dieses Kriterium zu Woche 25 im Vergleich zu 15 % des Placebo-Arms. Der Unterschied zwischen den Behandlungsarmen ist statistisch signifikant (RR: 2,65 (95%-KI: [1,21; 5,82])).

Responderanalysen für Verschlechterungen wurden weder in den Studienunterlagen prädefiniert noch post hoc für das Modul 4 durchgeführt. Aufgrund des Therapiefokus auf Verbesserungen und des nur langsam progredienten Erkrankungsverlaufs, bei einer Beobachtungsdauer von 24 Wochen sowie der Tatsache, dass während Teil 1 der Studie MOTION nur eine Person eine Krankheitsprogression erlitt, erscheint der Fokus auf Verbesserungen plausibel.

Zur Erfassung von Schmerzen wurden 2 Items des BPI-SF in der Studie MOTION täglich erhoben und berichtet: „Durchschnittlicher Schmerz“ und „Stärkster Schmerz“ innerhalb der letzten 24 Stunden. Hinsichtlich der Auswertung war eine Responderanalyse mit dem Kriterium einer 30%igen Verbesserung bezogen auf den Baseline-Wert unter Berücksichtigung der Schmerzmittelaufnahme vorgesehen. Aufgrund der Verteilung der Ausgangswerte ist davon auszugehen, dass für einen relevanten Anteil der Studienpopulation eine Verbesserung um 30 % des Ausgangswerts nicht einer Verbesserung um mindestens 15 % der Skalenspannweite von 10 Punkten entspricht. Folglich werden ausschließlich die stetigen Auswertungen in der Nutzenbewertung berichtet. Diese berücksichtigen zwar nicht potentielle Steigerungen bei der Einnahme narkotischer Analgetika, eine Erhöhung ist in der Studie MOTION jedoch lediglich für 2 Personen aus jedem



Studienarm dokumentiert und erscheint somit nicht relevant für die Ergebnisinterpretation. Für die in der Nutzenbewertung berücksichtigten, a priori definierten Veränderungen zu Woche 25 liegen jedoch keine Auswertungen für die Items des BPI-SF vor. Aus diesem Grund wird auf den Unterschied im Gesamteffekt der Behandlung Bezug genommen. Hier zeigen sich sowohl für „Stärkster Schmerz“ (LS-Mean-Differenz: -1,2 (95%-KI: [-1,8; -0,6]), Hedges' g: -0,79 (95%-KI: [-1,18; -0,40])) als auch für „Durchschnittlicher Schmerz“ (LS-Mean-Differenz: -1,0 (95%-KI: [-1,6; -0,5]), Hedges' g: -0,76 (95%-KI: [-1,15; -0,37])) jeweils statistisch signifikante Ergebnisse, für die auf Basis von Hedges' g ein klinisch relevanter Effekt plausibel erscheint.

Beim allgemeinen Gesundheitszustand mittels EQ-5D-VAS zeigte sich anhand der post hoc durchgeführten Responderanalyse für Verbesserungen zu Woche 25 ein statistisch signifikanter Effekt zugunsten von Vimseltinib (RR: 3,20 (95%-KI: [1,21; 8,44])). Das Responsekriterium einer Verbesserung um mindestens 15 Punkte im Vergleich zu Baseline (15 % der Skalenspannweite) wurde von 31 % der Personen des Vimseltinib-Arms und von 10 % des Placebo-Arms erfüllt.

Für die allgemeine Krankheitssymptomatik anhand des PGIC zeigte sich zu Woche 25 ein statistisch signifikanter Effekt zugunsten von Vimseltinib (RR: 2,95 (95%-KI: [1,63; 5,35])), wobei Personen mit jeglicher Verbesserung als Responder gewertet wurden.

Insgesamt resultiert aus den Ergebnissen für die Morbidität ein Vorteil für Vimseltinib im Vergleich zu Placebo für die Ergebnisse zu Woche 25, der konsistent über alle Endpunkte hinweg besteht. Das Verzerrungspotential für alle Ergebnisse wird jedoch aufgrund einer potentiellen Entblindung durch das spezifische Nebenwirkungsprofil von Vimseltinib im Interventionsarm (siehe Endpunkt „Unerwünschte Ereignisse“) und des hohen Anteils an fehlenden Werten als hoch bewertet. Allerdings sind die beobachteten Effekte bei den Endpunkten sehr groß, sodass die Ergebnissicherheit bezüglich des Vorhandenseins eines Effekts nicht eingeschränkt ist.

## 4.6 Lebensqualität

Aufgrund der fehlenden Erhebung der Lebensqualität in der Studie MOTION können keine Aussagen zum Einfluss von Vimseltinib auf die Lebensqualität getroffen werden.

## 4.7 Sicherheit

Alle UE wurden ab dem Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis zum Ende der Studienteilnahme (letzte Studienvisite) kontinuierlich erhoben. Für die Auswertung wurden alle UE berücksichtigt, die ab Erhalt der ersten Dosis der Studienmedikation bis 30 Tage nach letztmaligem Erhalt der Studienmedikation auftraten. Die mediane Behandlungsdauer war mit 5,5 Monaten in beiden Studienarmen gleich. Die mediane Studiendauer ist in Modul 4 mit 24 Wochen in jedem Behandlungsarm angegeben.

Die relativen Risiken (RR) wurden post hoc berechnet und es liegen widersprüchliche Angaben bezüglich der Berücksichtigung der Stratifizierungsfaktoren in der statistischen Analyse zwischen Modul 4 und den Studiendokumenten vor. Es wird davon ausgegangen, dass die Analysen unstratifiziert erfolgten.

Obwohl der pU keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vorlegt, ist basierend auf den aufgetretenen UE nicht davon auszugehen, dass Ereignisse der Grunderkrankung in signifikantem Umfang mit in die Erfassung der UE einfließen. Symptome der untersuchten Krankheit durften in der Studie MOTION nicht als UE erfasst werden und eine Krankheitsprogression ist lediglich für 1 Person dokumentiert. In beiden Behandlungsgruppen gingen knapp 90 % der Personen in Studienteil 2 über, sodass die 30-tägige Sicherheitsnachbeobachtung offensichtlich nur für einen



kleinen Anteil mit in die vorliegenden Sicherheitsauswertungen einging. Da der Anteil in beiden Studienarmen jedoch vergleichbar ist, sind daraus keine verzerrenden Aspekte zu erwarten.

Während der 25-wöchigen randomisierten Behandlungsphase in Teil 1 der Studie MOTION erlitt jede Person aus dem Interventionsarm mindestens ein behandlungsbedingtes UE, verglichen mit 95 % aus der Placebo-Gruppe. Der Unterschied bei schweren UE (CTCAE-Grad  $\geq 3$ ) war statistisch signifikant zuungunsten von Vimseltinib (RR: 3,64 (95%-KI: [1,38; 9,60])). Auch SUE (7 vs. 3 %) und UE, die zum Abbruch der Studienmedikation führten (6 vs. 0 %), traten numerisch häufiger unter Vimseltinib auf, die Unterschiede waren jedoch nicht statistisch signifikant.

Hinsichtlich der UE nach SOC und PT traten die SOC „Augenerkrankungen (74 vs. 23 %)“, „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort (77 vs. 46 %)“ sowie „Erkrankungen der Haut und des Unterhautgewebes (58 vs. 15 %)“ am häufigsten auf. Die Unterschiede zwischen den Behandlungsgruppen waren jeweils statistisch signifikant zuungunsten von Vimseltinib (RR: 3,19 (95%-KI: [1,77; 5,73])); RR: 1,67 (95%-KI: [1,17; 2,39]); RR: 3,76 (95%-KI: [1,76; 8,03])). Weiterhin zeigte sich ein statistisch signifikanter Effekt zuungunsten von Vimseltinib für die SOC „Untersuchungen“ (RR: 4,70 (95%-KI: [1,81; 12,2])). Auch auf Ebene der PT liegen ausschließlich statistisch signifikante Unterschiede zuungunsten von Vimseltinib bei folgenden 5 PT vor: „Periorbitalödem“ (RR: 3,48 (95%-KI: [1,48; 8,16])), „Gesichtsödem“ (RR: 4,07 (95%-KI: [1,31; 12,6])), „Aspartat-Aminotransferase erhöht“ (RR: 8,93 (95%-KI: [1,24; 64,3])), „Kreatinphosphokinase im Blut erhöht“ (RR: 19,5 (95%-KI: [1,21; 314,7])) und „Pruritus“ (RR: 3,76 (95%-KI: [1,20; 11,7])).

Schwere UE bei  $\geq 5$  % der Patientinnen und Patienten in mindestens einem Studienarm sind lediglich für die SOC „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“, „Infektionen und parasitäre Erkrankungen“ und „Untersuchungen“ sowie das PT „Kreatinphosphokinase im Blut erhöht“ dokumentiert. Abgesehen von der SOC „Allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ traten diese Ereignisse nur im Vimseltinib-Arm auf. Die Unterschiede waren jedoch jeweils nicht statistisch signifikant.

Im Rahmen der randomisierten Phase der Studie MOTION traten keine SUE auf Ebene der SOC und PT auf, welche bei mindestens 5 % der Patientinnen und Patienten in einem Studienarm aufgetreten sind. UE von besonderem Interesse waren für die Studie MOTION nicht definiert.

Aufgrund spezifischer charakteristischer UE im Interventionsarm, wie beispielsweise Periorbitalödem (45 %), Gesichtsödem (31 %), Pruritus (29 %) und Ausschlag (19 %), kann eine potentielle Entblindung der Studienteilnehmenden nicht ausgeschlossen werden. Zusammen mit der unbegründet fehlenden Stratifikation der Analysen ist von einem hohen Verzerrungspotential für Abbrüche aufgrund von UE sowie UE jeglichen Schweregrads nach SOC und PT auszugehen. Für schwere UE und SUE resultiert aus der fehlenden Stratifikation ein unklares Verzerrungspotential. Aufgrund der Effektschätzer und der p-Werte für die Gesamtraten der SUE und der schweren UE ist dennoch nicht von einer eingeschränkten Ergebnissicherheit auszugehen.

Insgesamt ist die Aussagekraft der Ergebnisse zu den UE aufgrund der relativ kurzen randomisierten Behandlungsphase von 24 Wochen limitiert. Auch die „European Medicines Agency“ (EMA) erachtet das Sicherheitsprofil von Vimseltinib aufgrund einer limitierten Expositionszeit sowie der begrenzten Patientenzahl als inkomplett charakterisiert [8]. Dies zeigt sich in relevant ansteigenden Anteilen von SUE und schweren UE bei längerfristiger Beobachtung. Die EMA sieht mehrere Sicherheitssignale hinsichtlich Leber-, Bluthochdruck- sowie Nieren-assoziiierter UE und anhaltender Hauttoxizitäten mit einer Vielzahl an UE-bedingten Dosismodifikationen.

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

Die Angaben der Fachinformation sind zu berücksichtigen.

## 6 Zusammenfassung der Nutzenbewertung

Vimseltinib ist zugelassen für erwachsene Patientinnen und Patienten mit symptomatischen tenosynovialen Riesenzelltumoren (TGCT), die mit einer klinisch relevanten Verschlechterung der körperlichen Funktionsfähigkeit assoziiert sind und bei denen chirurgische Optionen ausgeschöpft sind oder zu einer inakzeptablen Morbidität oder Behinderung führen würden. Die Nutzenbewertung von Vimseltinib basiert auf der zulassungsbegründenden Studie MOTION, einer randomisierten, multizentrischen, Placebo-kontrollierten Studie zur Untersuchung der Wirksamkeit und Sicherheit von Vimseltinib bei Personen mit TGCT bestehend aus einem doppelblinden Teil 1 und einem offenen, einarmigen Teil 2.

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 Vimseltinib 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 21: Zusammenfassende Darstellung der Ergebnisse der Studie MOTION*

Studie MOTION Darstellung der Ergebnisse	Vimseltinib N = 83			Placebo N = 39			Wirkstoff vs. Komparator	Effekt
Mortalität								
Todesfälle <sup>1)</sup>	N	Personen mit Ereignis, n (%)		N	Personen mit Ereignis, n (%)		Effektschätzer	
	83	0		40	0		-	↔
Morbidität								
PROMIS PF <sup>2)</sup>	N <sup>3)</sup>	Baseline MW (SD)	Änderung Woche 25 zu Baseline LS Mean (SE)	N <sup>3)</sup>	Baseline MW (SD)	Änderung Woche 25 zu Baseline LS Mean (SE)	LS-Mean- Differenz [95%-KI] <sup>4)</sup> ; p-Wert	
Veränderung zu Woche 25	63	39,0 (6,14)	4,6 (0,96)	30	38,5 (5,98)	1,3 (0,88)	3,3 [1,4; 5,2]; 0,0007 <sup>5)</sup>	↑
Steifigkeit mittels NRS <sup>6)</sup>	N	Personen mit Ereignis, n (%)		N	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert <sup>7)</sup>	
Verbesserung um ≥ 2 Punkte zu Woche 25	83	32 (38,6)		40	6 (15,0)		2,65 [1,21; 5,82]; 0,01	↑

Studie MOTION Darstellung der Ergebnisse	Vimseltinib N = 83			Placebo N = 39			Wirkstoff vs. Komparator	Effekt
Schmerzen mittels BPI-SF <sup>8)</sup>	N <sup>3)</sup>	Baseline MW (SD)	Änderung Woche 25 zu Baseline LS Mean (SE)	N <sup>3)</sup>	Baseline MW (SD)	Änderung Woche 25 zu Baseline LS Mean (SE)	LS-Mean- Differenz [95%-KI] <sup>4)</sup> ; p-Wert	
Stärkster Schmerz	68	5,5 (2,11)	-2,6 (0,33)	31	6,0 (1,78)	-1,0 (0,39)	-1,2 [-1,8; -0,6]; < 0,0001 <sup>9)</sup>	↑
Durchschnittlicher Schmerz	68	4,62 (2,00)	-2,4 (0,30)	31	5,19 (1,96)	-1,0 (0,36)	-1,0 [-1,6; -0,5]; 0,0001 <sup>9)</sup>	↑
EQ-5D-VAS <sup>10)</sup>	N	Personen mit Ereignis, n (%)		N	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert	
Verbesserung um ≥ 15 Punkte zu Woche 25	83	26 (31,3)		40	4 (10,0)		3,20 [1,21; 8,44]; 0,02	↑
PGIC- Krankheits- symptomatik	N	Personen mit Ereignis, n (%)		N	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert	
Verbesserung <sup>11)</sup>	83	55 (66,3)		40	9 (22,5)		2,95 [1,63; 5,35]; 0,0004	↑
Sicherheit <sup>12)</sup>								
Unerwünschte Ereignisse	N <sup>13)</sup>	Personen mit Ereignis, n (%)		N <sup>13)</sup>	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert <sup>14)</sup>	
Schwere UE (CTCAE-Grad≥3)	83	31 (37,3)		39	4 (10,3)		3,64 [1,38; 9,60]; 0,01	↓
SUE	83	6 (7,2)		39	1 (2,6)		2,82 [0,35; 22,62]; 0,33	↔
UE, das zum Abbruch der Studienmedi- kation führte	83	5 (6,0)		39	0 (0,0)		5,24 [0,30; 92,4]; 0,26	↔

<sup>1)</sup> Der Endpunkt wurde post hoc als separater Endpunkt für das Dossier Modul 4 ausgewertet.

<sup>2)</sup> T-transformierte Werte, MW = 50, SD = 10. Höhere Werte bedeuten eine bessere körperliche Funktionsfähigkeit.

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

<sup>4)</sup> MMRM unter Verwendung des Sandwich-Schätzers zur Schätzung der Varianz-Kovarianz-Matrix. Die abhängige Variable war dabei die Veränderung gegenüber dem Ausgangswert. Jedes dieser Modelle enthielt feste Effekte für die Behandlungsgruppe, den Zeitpunkt sowie die Interaktion zwischen Behandlungsgruppe und Zeitpunkt. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorklassifikation“ und „Region“. Zudem wurde für den Ausgangswert des entsprechenden Endpunkts als Co-Faktor adjustiert.

<sup>5)</sup> Es liegt kein Hedges'g für die Differenz der Veränderung zu Woche 25 vor. Auf Basis der Ergebnisse des Hedges'g für den Gesamteffekt bis Woche 25 (siehe Tabelle 15), erscheint das Vorliegen eines klinisch relevanten Effekts auch für die Differenz der Veränderung zu Woche 25 plausibel.

<sup>6)</sup> Werte von 0 bis 10; höhere Werte bedeuten eine stärkere Steifigkeit.

<sup>7)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Werts wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorklassifikation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

- <sup>8)</sup> Es wurden ausschließlich die Items „Stärkster Schmerz“ und „Durchschnittlicher Schmerz“ erhoben. Skala von 0 bis 10; niedrigere Werte bedeuten weniger Schmerzen.
- <sup>9)</sup> Es liegen keine Ergebnisse für die Differenz der Veränderung zu Woche 25 vor. Auf Basis der Ergebnisse des Hedges'g für den Gesamteffekt bis Woche 25, erscheint das Vorliegen eines klinisch relevanten Effekts auch für die Differenz der Veränderung zu Woche 25 plausibel (siehe Tabelle 15).
- <sup>10)</sup> Werte von 0 bis 100; höhere Werte entsprechen dem bestmöglichen Gesundheitszustand
- <sup>11)</sup> Eine Verbesserung im PGIC war definiert als „sehr viel besser“, „viel besser“ und „wenig besser“.
- <sup>12)</sup> Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind in Kapitel 3.5 dargestellt.
- <sup>13)</sup> Die Sicherheitspopulation umfasst alle Personen, welche zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben.
- <sup>14)</sup> Log-Binomial-Regression mit dem Term „Behandlung“.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; CTCAE: Common Terminology Criteria for Adverse Events; EQ-5D-VAS: Visuelle Analogskala des European Quality of Life 5-Dimension; KI: Konfidenzintervall; LS: Least Squares; MMRM: Mixed Model for Repeated Measures; MW: Mittelwert; NRS: Numerische Bewertungsskala; PGIC: Patient Global Impression of Change; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; RR: Relatives Risiko; SD: Standardabweichung; SE: Standardfehler; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## Referenzen

1. **Bernthal NM, Spierenburg G, Healey JH, Palmerini E, Bauer S, Topp Study Group, et al.** The diffuse-type tenosynovial giant cell tumor (dt-TGCT) patient journey: a prospective multicenter study. *Orphanet J Rare Dis* 2021;16(1):191.
2. **Byrne MJ, Nowak AK.** Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann Oncol* 2004;15(2):257-260.
3. **Centers for Disease Control and Prevention (CDC).** Calculating total daily dose of opioids for safer dosage [online]. CDC; 2016. [Zugriff: 22.01.2026]. URL: <https://stacks.cdc.gov/view/cdc/38481>.
4. **Deciphera Pharmaceuticals.** A phase 3, randomized, placebo-controlled, double-blind study of vimseltinib to assess the efficacy and safety in patients with tenosynovial giant cell tumor (MOTION); study number DCC-3014-03-001; clinical study report [unveröffentlicht]. 22.03.2024.
5. **Deciphera Pharmaceuticals (Netherlands).** Romvimza 14 mg/ -20 mg/ -30 mg Hartkapseln [online]. Stand: 09.2025. Frankfurt/Main. [Zugriff: 05.11.2025]. URL: <https://www.fachinfo.de/fi/detail/025500/romvimza-14-mg-20-mg-30-mg-hartkapseln>.
6. **Deciphera Pharmaceuticals (Netherlands).** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Vimseltinib (Romvimza), Behandlung von erwachsenen Patienten mit symptomatischen tenosynovialen Riesenzelltumoren (Tenosynovial Giant Cell Tumours, TGCT), die mit einer klinisch relevanten Verschlechterung der körperlichen Funktionsfähigkeit assoziiert sind und bei denen chirurgische Optionen ausgeschöpft sind oder zu einer inakzeptablen Morbidität oder Behinderung führen würden; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 30.10.2025.
7. **Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al.** New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45(2):228-247.
8. **European Medicines Agency (EMA).** Romvimza: EPAR public assessment report EMEA/H/C/006363/0000 [online]. Last update: 24.07.2025. Amsterdam (NED): EMA; 2025. [Zugriff: 05.11.2025]. URL: [https://www.ema.europa.eu/en/documents/assessment-report/romvimza-epar-public-assessment-report\\_en.pdf](https://www.ema.europa.eu/en/documents/assessment-report/romvimza-epar-public-assessment-report_en.pdf).
9. **Evidera.** Exit interviews with patients with tenosynovial giant cell tumor (TGCT), DCC-3014-03-001: final qualitative study report [unveröffentlicht]. 2023.
10. **Gelderblom H, Bhadri V, Stacchiotti S, Bauer S, Wagner AJ, van de Sande M, et al.** Vimseltinib versus placebo for tenosynovial giant cell tumour (MOTION): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* 2024;403(10445):2709–2719.
11. **Gelhorn HL, Tong S, McQuarrie K, Vernon C, Hanlon J, Maclaine G, et al.** Patient-reported symptoms of tenosynovial giant cell tumors. *Clin Ther* 2016;38(4):778–793.
12. **Gelhorn HL, Ye X, Speck RM, Tong S, Healey JH, Bukata SV, et al.** The measurement of physical functioning among patients with tenosynovial giant cell tumor (TGCT) using the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Patient Rep Outcomes* 2019;3(1):6.

13. **HealthMeasures.** PROMIS physical function: user manual and scoring instructions [online]. Evanston (USA): Northwestern University; 2025. [Zugriff: 22.01.2026]. URL: [https://www.healthmeasures.net/images/PROMIS/manuals/Scoring\\_Manual\\_Only/PRO\\_MIS\\_Physical\\_Function\\_User\\_Manual\\_and\\_Scoring\\_Instructions\\_24Nov2025.pdf](https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PRO_MIS_Physical_Function_User_Manual_and_Scoring_Instructions_24Nov2025.pdf).
14. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).** Cipaglusidase alfa (Morbus Pompe), Addendum zum Projekt A23-79 (Dossierbewertung); Projekt: A23-133 [online]. Köln (GER): IQWiG; 2024. [Zugriff: 22.01.2026]. (IQWiG-Berichte Nr. 1700). URL: [https://www.iqwig.de/download/a23-133\\_cipaglusidase-alfa\\_addendum-zum-projekt-a23-79\\_v1-0.pdf](https://www.iqwig.de/download/a23-133_cipaglusidase-alfa_addendum-zum-projekt-a23-79_v1-0.pdf).
15. **Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.** Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-766 e761.
16. **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.
17. **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.
18. **Palmerini E, Staals EL, Maki RG, Pengo S, Cioffi A, Gambarotti M, et al.** Tenosynovial giant cell tumour/pigmented villonodular synovitis: outcome of 294 patients before the era of kinase inhibitors. *Eur J Cancer* 2015;51(2):210-217.
19. **Speck RM, Ye X, Bernthal NM, Gelhorn HL.** Psychometric properties of a custom Patient-Reported Outcomes Measurement Information System (PROMIS) physical function short form and worst stiffness numeric rating scale in tenosynovial giant cell tumors. *J Patient Rep Outcomes* 2020;4(1):61.
20. **Stacchiotti S, Durr HR, Schaefer IM, Woertler K, Haas R, Trama A, et al.** Best clinical management of tenosynovial giant cell tumour (TGCT): a consensus paper from the community of experts. *Cancer Treat Rev* 2023;112:102491.
21. **Tap WD, Wainberg ZA, Anthony SP, Ibrahim PN, Zhang C, Healey JH, et al.** Structure-guided blockade of CSF1R kinase in tenosynovial giant-cell tumor. *N Engl J Med* 2015;373(5):428-437.
22. **Von Korff M, Saunders K, Thomas Ray G, Boudreau D, Campbell C, Merrill J, et al.** De facto long-term opioid therapy for noncancer pain. *Clin J Pain* 2008;24(6):521-527.
23. **World Health Organization (WHO).** WHO guidelines for the pharmacological and radiotherapeutic management of cancer pain in adults and adolescents [online]. Genf (SUI): WHO; 2018. [Zugriff: 22.01.2026]. URL: <https://iris.who.int/server/api/core/bitstreams/809a021b-e75c-4f9c-b002-6fb13bf1a0f5/content>.



## Anhang

### Bewegungsumfang

*Tabelle 22: Bewegungsumfang, MMRM-Analysen anhand beobachteter Werte zu Woche 25;  
Studie MOTION, ITT-Population*

<b>Studie MOTION Bewegungsumfang<sup>1)</sup></b>	<b>Vimseltinib N = 83</b>	<b>Placebo N = 40</b>
<i>Baseline</i>		
Verfügbare Werte, n (%)	79 (95,2)	38 (95,0)
MW (SD)	63,0 (29,4)	62,9 (32,2)
<i>Woche 25</i>		
Verfügbare Werte, n (%)	73 (88,0)	33 (82,5)
MW (SD)	83,6 (28,1)	68,3 (35,3)
<i>Veränderung zu Woche 25</i>		
Verfügbare Werte, n (%)	73 (88,0)	33 (82,5)
LS Mean (SE) <sup>2)</sup>	18,4 (6,46)	3,8 (7,19)
<i>Gesamteffekt der Behandlung bis Woche 25</i>		
Verfügbare Werte, n (%)	77 (92,8)	36 (90,0)
LS Mean (SE) <sup>2)</sup>	17,8 (6,64)	3,9 (6,94)
<i>Veränderung im Gesamteffekt</i>		
LS-Mean-Differenz [95%-KI] <sup>2)</sup> ; p-Wert	13,9 [4,1; 23,7]; 0,006	
<i>Veränderung zu Woche 25</i>		
LS-Mean-Differenz [95%-KI] <sup>2)</sup> ; p-Wert	14,6 [4,0; 25,3]; 0,008	

<sup>1)</sup> Aktiver Bewegungsumfang. Werte in Prozent relativ zu den AMA-Referenzwerten; niedrigere Werte bedeuten eine größere Abweichung vom Referenzstandard.

<sup>2)</sup> MMRM unter Verwendung einer unstrukturierten Kovarianzmatrix mit den Termen „Behandlung“, „Besuch“, „Region“, „Gelenktyp“, „Ausgangswert“ und „Wechselwirkung zwischen Behandlung und Besuch“. MMRM unter Verwendung des Sandwich-Schätzers zur Schätzung der Varianz-Kovarianz-Matrix.

Abkürzungen: AMA: American Medical Association; ITT: Intention to Treat; KI: Konfidenzintervall; LS: Least Squares; MMRM: Mixed Model for Repeated Measures; MW: Mittelwert; SD: Standardabweichung; SE: Standardfehler.

**PGIC**

*Tabelle 23: Anteil der Personen mit einer Verbesserung im PGIC zu Woche 25;  
Studie MOTION, ITT-Population*

Studie MOTION PGIC Responderanalyse zu Woche 25	Vimseltinib N = 83	Placebo N = 40
Verbesserung im PGIC-ROM <sup>1)</sup>		
Verfügbare Werte, n (%)	64 (77,1)	30 (75,0)
Responder, n (%)	53 (63,9)	8 (20,0)
RR [95%-KI]; p-Wert <sup>2)</sup>	3,24 [1,71; 6,13]; 0,0003	
Verbesserung im PGIC-PF <sup>1)</sup>		
Verfügbare Werte, n (%)	64 (77,1)	30 (75,0)
Responder, n (%)	55 (66,3)	9 (22,5)
RR [95%-KI]; p-Wert <sup>2)</sup>	2,95 [1,63; 5,35]; 0,0004	

<sup>1)</sup> Verbesserung war definiert als „sehr viel besser“, „viel besser“ und „wenig besser“.

<sup>2)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Wertes wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorlokalisation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

Abkürzungen: ITT: Intention to Treat; KI: Konfidenzintervall; PF: Körperliche Funktionsfähigkeit; PGIC: Patient Global Impression of Change; ROM: Bewegungsumfang; RR: Relatives Risiko.

## Tumoransprechen

Tabelle 24: Anteil der Personen mit Tumoransprechen zu Woche 25; Studie MOTION, ITT-Population

Studie MOTION Tumoransprechen Responderanalyse zu Woche 25	Vimseltinib N = 83	Placebo N = 40
Tumoransprechen gemäß IRR nach RECIST <sup>1)</sup>		
Verfügbare Werte, n (%)	75 (90,4)	33 (82,5)
Nicht evaluierbar n (%)	8 (9,6)	7 (17,5)
Responder, n (%)	33 (39,8)	0
RR [95%-KI]; p-Wert <sup>2)</sup>	32,7 [2,06; 520,5] 0,01	
Tumoransprechen gemäß Prüfpersonal nach RECIST <sup>1)</sup>		
Verfügbare Werte, n (%)	76 (91,6)	36 (90,0)
Nicht evaluierbar n (%)	7 (8,4)	4 (10,0)
Responder, n (%)	28 (33,7)	1 (2,5)
RR [95%-KI]; p-Wert <sup>2)</sup>	13,5 [1,91; 95,6]; 0,0092	
Tumoransprechen gemäß IRR nach mRECIST <sup>1)</sup>		
Verfügbare Werte, n (%)	75 (90,4)	35 (87,5)
Nicht evaluierbar n (%)	8 (9,6)	5 (12,5)
Responder, n (%)	50 (60,2)	1 (2,5)
RR [95%-KI]; p-Wert <sup>2)</sup>	24,2 [3,47; 168,9]; 0,0013	

<sup>1)</sup> Definiert als CR oder PR.

<sup>2)</sup> Zur Schätzung des RR, des entsprechenden 95%-Wald-KI und des p-Wertes wurde eine log-binomiale Regression verwendet. Die Analysen erfolgten stratifiziert nach den Stratifizierungsfaktoren „Tumorlokalisation“ und „Region“. Fehlende Werte nach Baseline wurden mittels Non-Responder-Imputation ersetzt.

Abkürzungen: CR: Komplettes Ansprechen; IRR: Unabhängige radiologische Überprüfung; ITT: Intention to Treat; KI: Konfidenzintervall; mRECIST: modified Response Evaluation Criteria In Solid Tumors; PR: Partielles Ansprechen; RECIST: Response Evaluation Criteria In Solid Tumors; RR: Relatives Risiko.

Tabelle 25: Faktoren des Gesamtansprechens gemäß RECIST (Version 1.1)

Kategorie	Definition
CR	Verschwinden aller Zielläsionen. Alle pathologischen Lymphknoten müssen in der kurzen Achse kleiner als 10 mm sein. Nicht-noduläre Zielläsionen müssen verschwunden sein.
PR	Mindestens 30 % Abnahme der Summe des Durchmessers der Zielläsionen, wobei die Summe der Ausgangsdurchmesser als Referenz dient.
PD	Mindestens 20 % Zunahme der Summe des Durchmessers der Zielläsionen, wobei die kleinste Summe in der Studie als Referenz dient. Zusätzlich zu der relativen Zunahme von 20 % muss die Summe auch eine absolute Zunahme von mind. 5 mm aufweisen.
SD	Weder ausreichende Abnahme der Summe des Durchmessers der Zielläsionen, um sich für PR zu qualifizieren, noch ausreichende Zunahme, um sich für PD zu qualifizieren. Dabei dient die kleinste Summe des Durchmessers während der Studie als Referenz.
Nicht bewertbar	Ein Fortschreiten der Erkrankung wurde nicht dokumentiert, und eine oder mehrere Zielläsionen wurden nicht bewertet oder mit einer anderen Methode als der zu Baseline bewertet, was eine Vergleichbarkeit unmöglich macht.

Abkürzungen: CR: Komplettes Ansprechen; PD: Progrediente Erkrankung; PR: Partielles Ansprechen; RECIST: Response Evaluation Criteria In Solid Tumors; SD: Stabile Erkrankung.