

## 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: Nirogacestat

Datum der Veröffentlichung: 15. Januar 2026

Version 1.1

**Die vorliegende Version 1.1. vom 2. April 2026 ersetzt Version 1.0 der Nutzenbewertung vom 15. Januar 2026.**

Version 1.1. enthält im Vergleich zu Version 1.0 folgende Änderungen:

In Tabelle 8 (Seite 35) wurde die Zuordnung der Fußnote 3 korrigiert. Die Fußnote wird nun sachlich korrekt der Kategorie „ITT adäquat“ zugeordnet (zuvor: „Sonstige, das Verzerrungspotential beeinflussende Punkte“).

Diese Anpassung stellt eine rein formale Richtigstellung dar und hat keine Auswirkungen auf die Nutzenbewertung.

## Inhaltsverzeichnis

Tabellenverzeichnis .....	3
Abbildungsverzeichnis .....	5
Abkürzungsverzeichnis .....	7
Hintergrund.....	9
1 Fragestellung.....	10
2 Auswahl und Methodik der relevanten Studien .....	11
2.1 Studienbasis für die Nutzenbewertung .....	11
2.2 Beschreibung der eingeschlossenen Studie .....	11
2.3 Endpunkte .....	17
2.3.1 Mortalität.....	18
2.3.2 Morbidität.....	18
2.3.3 Lebensqualität .....	27
2.3.4 Sicherheit .....	28
2.3.5 Erhebungszeitpunkte .....	30
2.4 Statistische Methoden.....	31
2.5 Verzerrungspotential auf Studien- und Endpunktebene .....	34
3 Ergebnisse der eingeschlossenen Studie .....	36
3.1 Studiencharakteristika und Studienmedikation .....	36
3.2 Mortalität .....	43
3.3 Morbidität .....	43
3.4 Lebensqualität.....	56
3.5 Sicherheit .....	58
4 Diskussion der Methodik und Ergebnisse.....	67
4.1 Design und Methodik der Studie.....	67
4.2 Zulassungsstatus und Zulassungspopulation.....	68
4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit.....	69
4.4 Mortalität .....	69
4.5 Morbidität .....	69
4.6 Lebensqualität.....	71
4.7 Sicherheit .....	71
5 Anforderung an eine qualitätsgesicherte Anwendung .....	73
6 Zusammenfassung der Nutzenbewertung .....	74
Referenzen .....	77
Anhang .....	79

## Tabellenverzeichnis

Tabelle 1: Übersicht über die Studienbasis .....	11
Tabelle 2: Charakterisierung der Studie DeFi .....	11
Tabelle 3: Für die Nutzenbewertung relevante Protokolländerungen der Studie DeFi .....	15
Tabelle 4: Charakterisierung der Intervention in der Studie DeFi.....	16
Tabelle 5: Zusammenfassung der Endpunktbewertung der Studie DeFi .....	17
Tabelle 6: Erhebungszeitpunkte der berücksichtigten und ergänzend dargestellten Endpunkte in der Studie DeFi .....	30
Tabelle 7: Verzerrungspotential der Studie DeFi .....	34
Tabelle 8: Verzerrungspotential der in der Nutzenbewertung berücksichtigten und ergänzend dargestellten Endpunkte der Studie DeFi.....	35
Tabelle 9: Allgemeine Angaben zur DB-Phase; Studie DeFi (Datenschnitt: 07.04.2022).....	36
Tabelle 10: Charakterisierung der Studienpopulation; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	37
Tabelle 11: Angaben zur Exposition mit der Studienmedikation; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022) .....	41
Tabelle 12: Begleitmedikation bei mindestens 15 % der Studienpopulation; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022) .....	42
Tabelle 13: Ergebnisse zum PFS (ergänzend dargestellt); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	44
Tabelle 14: GODDESS-DTSS – Responderanalyse (Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	45
Tabelle 15: GODDESS-DTIS – Responderanalyse (Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	47
Tabelle 16: GODDESS-DTIS – Subgruppenanalysen (Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	49
Tabelle 17: BPI-SF – Responderanalyse (Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	50
Tabelle 18: PGIC – Responderanalyse (Zeit bis zu jeglicher erstmaligen Verschlechterung); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	53
Tabelle 19: PGIC – Subgruppenanalyse (Zeit bis zu jeglicher Verschlechterung); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	54
Tabelle 20: EORTC QLQ-C30: Symptomskalen – Responderanalyse (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	55

Tabelle 21: EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus – Responderanalysen (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	57
Tabelle 22: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022) .....	58
Tabelle 23: UE mit Inzidenz $\geq 10\%$ in mindestens einem Studienarm nach Systemorganklasse und Preferred Term, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022) .....	59
Tabelle 24: Schwere UE (CTCAE-Grad $\geq 3$ ), die bei $\geq 5\%$ der Personen in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022) .....	63
Tabelle 25: SUE, die bei $\geq 5\%$ der Personen in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022) .....	64
Tabelle 26: UESI; Studie DeFi (DB-Phase) Sicherheitspopulation (Datenschnitt: 07.04.2022).....	65
Tabelle 27: Zusammenfassende Darstellung der Ergebnisse der Studie DeFi (DB-Phase).....	74

## Abbildungsverzeichnis

Abbildung 1: Kaplan-Meier-Kurve für PFS (Zeit bis zu jeglicher qualifizierten klinischen Progression oder zum Tod); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	44
Abbildung 2: Kaplan-Meier-Kurve für GODDESS-DTSS – Gewichteter DTSS-Gesamtscore, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	46
Abbildung 3: Kaplan-Meier-Kurve für GODDESS-DTSS – Domänenscore „Intra-abdominale Symptome“, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalen- spannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	46
Abbildung 4: Kaplan-Meier-Kurve für GODDESS-DTIS – Domänenscore „Körperliche Funktion“, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	48
Abbildung 5: Kaplan-Meier-Kurve für GODDESS-DTIS – Domänenscore „Schlaf“, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	48
Abbildung 6: Kaplan-Meier-Kurve für GODDESS-DTIS – Domänenscore „Emotion“, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	49
Abbildung 7: Kaplan-Meier-Kurve für BPI-SF – Skala „Beeinträchtigung durch den Schmerz“, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	51
Abbildung 8: Kaplan-Meier-Kurve für BPI-SF – Skala „Schmerzintensität“, Zeit bis zur ersten Verschlechterung um $\geq 15\%$ der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	52
Abbildung 9: Kaplan-Meier-Kurve für den PGIC (allgemeiner Gesundheitszustand), Zeit bis zu jeglicher erstmaligen Verschlechterung, Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	53
Abbildung 10: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Fatigue“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	79
Abbildung 11: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Übelkeit und Erbrechen“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	79
Abbildung 12: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Schmerz“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	80
Abbildung 13: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Dyspnoe“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	80
Abbildung 14: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Schlaflosigkeit“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022).....	81

Abbildung 15: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Appetitverlust“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	81
Abbildung 16: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Obstipation“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	82
Abbildung 17: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Diarrhö“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	82
Abbildung 18: Kaplan-Meier-Kurve für EORTC QLQ-C30: Globaler Gesundheitsstatus (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	83
Abbildung 19: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Kognitive Funktion“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	83
Abbildung 20: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Emotionale Funktion“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	84
Abbildung 21: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Körperliche Funktion“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	84
Abbildung 22: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Rollenfunktion“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	85
Abbildung 23: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Soziale Funktion“ (Zeit bis zur ersten Verschlechterung um $\geq 10$ Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	85
Abbildung 24: Kaplan-Meier-Kurve für schwere UE (CTCAE-Grad $\geq 3$ ), Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) ...	86
Abbildung 25: Kaplan-Meier-Kurve für SUE, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	86
Abbildung 26: Kaplan-Meier-Kurve UE, die zum Abbruch der Studienmedikation führten, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022) .....	87

## Abkürzungsverzeichnis

AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
APC	Adenomatous Polyposis Coli
API	Durchschnittliche Schmerzintensität (Average Pain Intensity)
BPI-SF	Brief Pain Inventory – Short Form
CR	Vollständiges Ansprechen (Complete Response)
CTCAE	Common Terminology Criteria for Adverse Events
CTNNB1	Catenin Beta-1
DB	doppelblind
DTIS	Desmoid Tumor Impact Scale
DTSS	Desmoid Tumor Symptom Scale
EAC	Endpoint Adjudication Committee
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire
EPAR	European Public Assessment Report
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GODDESS	GOunder/DTRF DEsmoid Symptom/Impact Scale
HR	Hazard Ratio
ICC	Intraklassen-Korrelationskoeffizient
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to Treat
KI	Konfidenzintervall
MedDRA	Medical Dictionary for Regulatory Activities
MRT	Magnetresonanztomographie
N	Anzahl
NRS	Numerische Bewertungsskala (Numeric Rating Scale)
OLE	Open-Label-Extension
ORR	Gesamtansprechrate (Objective Response Rate)
PFS	Progressionsfreies Überleben (Progression-Free Survival)
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PR	Partielles Ansprechen (Partial Response)
PRO	Patientenberichteter Endpunkt (Patient-Reported Outcome)
PROMIS PF	Patient-Reported Outcomes Measurement Information System Physical Function
PT	Preferred Term
pU	pharmazeutischer Unternehmer
RECIST	Response Evaluation Criteria In Solid Tumors
RR	Relatives Risiko
SAP	Statistischer Analyseplan
SGB	Sozialgesetzbuch

SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis
UE	Unerwünschtes Ereignis
UESI	Unerwünschtes Ereignis von besonderem Interesse
VerfO	Verfahrensordnung des G-BA

## Hintergrund

Nirogacestat 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 Nirogacestat 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 Nirogacestat in seiner Sitzung am 13. Januar 2026 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 14. 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 15. Januar 2025 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

## 1 Fragestellung

Nirogacestat (Ogsiveo®) 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:

Behandlung von erwachsenen Patienten mit fortschreitenden Desmoidtumoren als Monotherapie, die eine systemische Behandlung erfordern.

Gemäß Fachinformation beträgt die zugelassene Dosierung 150 mg, die zweimal täglich morgens und abends oral eingenommen wird. Die Behandlung soll bis zum Fortschreiten der Erkrankung oder bis zu einer inakzeptablen Toxizität fortgeführt werden. [17]

## 2 Auswahl und Methodik der relevanten Studien

### 2.1 Studienbasis für die Nutzenbewertung

Tabelle 1: Übersicht über die Studienbasis

Studiename (Nummer)	Studie mit Dossier vom pU eingereicht	Studie vom pU als relevant für die Ableitung des Zusatznutzens erachtet	Studie relevant für die Nutzenbewertung	Ausschlussgrund
<b>Studien zum Nirogacestat</b>				
DeFi (NIR-DT-301) <sup>1)</sup>	Ja	Ja	Ja	-

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

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

#### Zur Nutzenbewertung für Nirogacestat herangezogene Unterlagen

- Dossier des pU zu Nirogacestat [13,14,15,16]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [2]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan (SAP) der Studie DeFi (NIR-DT-301) [18,19,20]
- Fachinformation zu Nirogacestat [17]
- Patientenberichtete Endpunkte (PRO) – Report und SAP [12,21]

### 2.2 Beschreibung der eingeschlossenen Studie

Die Nutzenbewertung von Nirogacestat im vorliegenden Anwendungsgebiet basiert auf der doppelblinden Phase (DB-Phase) der pivotalen Zulassungsstudie DeFi (NIR-DT-301). Die Studie und die Intervention werden in Tabelle 2 und Tabelle 4 charakterisiert.

Tabelle 2: Charakterisierung der Studie DeFi

Charakteristikum	Beschreibung
<b>Design und Studienablauf</b>	<p><b>Design</b> Multizentrische, randomisierte, doppelblinde, placebokontrollierte Phase-III-Studie bei erwachsenen Personen mit fortschreitenden Desmoidtumoren (DT), die eine systemische Therapie erfordern. Die Randomisierung erfolgte stratifiziert nach Zieltumorlokalisation:</p> <ul style="list-style-type: none"> <li>• intra-abdominal<sup>1)</sup></li> <li>• extra-abdominal</li> </ul> <p>Die Studie beinhaltet zwei Phasen:</p> <ul style="list-style-type: none"> <li>• DB-Phase (finaler Datenschnitt: 07.04.2022)<sup>2)</sup></li> <li>• Optionale OLE-Phase (finale Datenbanksperre: 19.12.2024)</li> </ul> <p><b>Studienablauf</b></p> <ul style="list-style-type: none"> <li>• Screening-Phase (bis zu 28 Tage)</li> <li>• Randomisierung (Zyklus 1, Tag 1)</li> <li>• DB-Phase (28-Tage-Zyklus)</li> <li>• Follow-up (30 Tage nach letzter Dosis; nur für Personen, die nicht in OLE-Phase übergehen)</li> </ul>

Charakteristikum	Beschreibung
	<p><u>DB-Phase</u> Verbleib in DB-Phase bis:</p> <ul style="list-style-type: none"> <li>• Tod <u>oder</u></li> <li>• vorzeitiger Abbruch der Behandlung <u>oder</u></li> <li>• radiographischer Progress (gemäß RECIST v1.1) mittels zentraler Bildauswertung <u>oder</u></li> <li>• klinische Progression festgestellt durch das Prüfpersonal <u>oder</u></li> <li>• die Studie durch den Sponsor gestoppt wird <u>oder</u></li> <li>• die erforderliche Anzahl von 51 PFS-Ereignissen eingetreten und die primäre PFS-Analyse abgeschlossen ist.</li> </ul> <p>Im Fall, dass die PFS-Analyse abgeschlossen ist <u>oder</u> eine radiographische Progression festgestellt wird, erfolgt eine Entblindung zum Zeitpunkt der EOT-Visite und die Personen hatten die Möglichkeit die optionale OLE-Phase zu beginnen, wenn die präspezifizierten Einschlusskriterien für die OLE-Phase<sup>3)</sup> erfüllt sind. Demgegenüber erfolgt für Personen, bei denen eine klinische Progression durch das Prüfpersonal festgestellt wurde, keine Entblindung zum Zeitpunkt der EOT-Visite; für diese Personen bestand keine Möglichkeit in die OLE-Phase überzugehen.</p> <p><u>Optionale OLE-Phase</u> Die erste Visite der OLE-Phase erfolgte entweder am selben Tag oder innerhalb von 24 Stunden nach der EOT-Visite der DB-Phase. Vor der Entblindung mussten sämtliche EOT-Untersuchungen sowie alle laufenden UE und SUE vom verblindeten Prüfpersonal hinsichtlich ihrer Kausalität bewertet und im eCRF erfasst werden. In der OLE-Phase erhielten alle Personen 150 mg Nirogacestat zweimal täglich in 28-Tage-Zyklen.</p> <p>Verbleib in OLE bis</p> <ul style="list-style-type: none"> <li>• Tod <u>oder</u></li> <li>• Abbruch der Behandlung <u>oder</u></li> <li>• radiographischer Progress (gemäß RECIST v1.1) mittels zentraler Bildauswertung <u>oder</u></li> <li>• klinische Progression festgestellt durch das Prüfpersonal <u>oder</u></li> <li>• vorzeitige Beendigung der Studie durch den Sponsor <u>oder</u></li> <li>• Zulassung und Verfügbarkeit von Nirogacestat.</li> </ul>
<b>Population</b>	<p><b>Wesentliche Einschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Erwachsene Personen (<math>\geq 18</math> Jahre) mit histologisch bestätigtem DT (durch einen lokalen Pathologen vor der Einwilligung) mit einer Progression von <math>\geq 20\%</math>, gemessen nach RECIST v1.1, innerhalb von 12 Monaten vor der Screening-Visite.</li> <li>• Erkrankungsbild: <ul style="list-style-type: none"> <li>◦ Behandlungsnaive, messbare fortschreitende DT, die als nicht operabel – ohne das Risiko signifikanter Morbidität – angesehen werden <u>oder</u></li> <li>◦ rezidivierende, messbare fortschreitende DT nach mindestens einer Therapielinie <u>oder</u></li> <li>◦ refraktäre, messbare fortschreitende DT nach mindestens einer Therapielinie.</li> </ul> </li> <li>• DT ohne unmittelbares Progressionsrisiko.</li> <li>• Archiviertes/Neues Tumorgewebe zur Bestätigung verfügbar.</li> <li>• Vorherige DT-Therapie muss <math>\geq 28</math> Tage oder 5 Halbwertszeiten (je nachdem, was länger ist) vor Studienbeginn abgeschlossen sein; alle Toxizitäten müssen auf <math>\leq</math> Grad 1 oder Baseline zurückgegangen sein.</li> <li>• Stabile NSAID-Einnahme (seit <math>\geq 28</math> Tagen) bei nicht-DT-Indikation vor Progression.</li> <li>• ECOG-Status <math>\leq 2</math>.</li> <li>• Angemessene Organ- und Knochenmarkfunktion: <ul style="list-style-type: none"> <li>◦ Absolute Neutrophilenzahl <math>\geq 1500/\mu\text{l}</math></li> <li>◦ Thrombozyten <math>\geq 100.000/\mu\text{l}</math></li> <li>◦ Hämoglobin <math>\geq 9\text{ g/dl}</math></li> </ul> </li> </ul>

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> <li>o Gesamtbilirubin <math>\leq 1,5 \times \text{ULN}</math> (isoliertes Bilirubin <math>&gt; 1,5 \times \text{ULN}</math> ist akzeptabel, wenn Bilirubin fraktioniert ist und direktes Bilirubin <math>&lt; 35 \%</math>)</li> <li>o AST/ALT <math>\leq 2 \times \text{ULN}</math></li> <li>o Serumkreatinin <math>\leq 1,5 \times \text{ULN}</math> oder Kreatinin-Clearance <math>\geq 60 \text{ ml/min}</math>, wenn Kreatinin <math>&gt; 1,5 \times \text{ULN}</math></li> <li>• Keine absorptionsbeeinträchtigenden gastrointestinalen Erkrankungen</li> </ul> <p><b>Wesentliche Ausschlusskriterien</b></p> <ul style="list-style-type: none"> <li>• Bekanntes Malabsorptionssyndrom oder bestehende gastrointestinale Erkrankungen, die die Absorption von Nirogacestat beeinträchtigen können (z. B. Magenbypass, Lap-Band oder andere Magenoperationen, die die Absorption verändern würden); die Verabreichung von Nirogacestat über eine nasogastrische Sonde oder eine Magensonde ist nicht erlaubt.</li> <li>• Vorerkrankungen innerhalb von 6 Monaten nach Unterzeichnung der Einwilligungserklärung:             <ul style="list-style-type: none"> <li>o Klinisch signifikante Herzkrankheit (NYHA-Klasse III oder IV),</li> <li>o Myokardinfarkt,</li> <li>o schwere/instabile Angina,</li> <li>o koronare/periphere Bypass-Operation,</li> <li>o symptomatische kongestive Herzinsuffizienz,</li> <li>o zerebrovaskulärer Unfall,</li> <li>o transitorische ischämische Attacke oder</li> <li>o symptomatische Lungenembolie.</li> </ul> </li> <li>• Vorgeschichte zusätzlicher Risikofaktoren für Torsades-de-Pointes-Tachykardie (z. B. Herzinsuffizienz, Hypokaliämie, familiäre Vorgeschichte von Long-QT-Syndrom).</li> <li>• Innerhalb der letzten 5 Jahre zum Zeitpunkt der Einwilligung Lymphom, Leukämie oder eine andere Malignität, mit Ausnahme von lokal wiederkehrendem Krebs, der kurativ behandelt wurde (z. B. reseziertes Basalzell- oder Plattenepithelkarzinom der Haut, oberflächliches Blasenkarzinom, Karzinom in situ des Gebärmutterhalses oder der Brust), ohne Nachweis einer metastatischen Erkrankung seit 3 Jahren zum Zeitpunkt der Einwilligung.</li> <li>• Aktuelle oder chronische Vorgeschichte von Lebererkrankungen oder bekannte hepatische oder biliäre Anomalien (außer Gilbert-Syndrom oder asymptotische Gallensteine).</li> <li>• Therapie mit Gamma-Sekretase-Inhibitoren oder einer Anti-Notch-Antikörpertherapie.</li> <li>• Behandlung für DT, einschließlich Tyrosinkinase-Inhibitoren, NSAID (chronische tägliche Anwendung – außer wie in Einschlusskriterium „Stabile NSAID-Einnahme (seit <math>\geq 28</math> Tagen) bei nicht-DT-Indikation vor Progression“) oder eine andere experimentelle Behandlung 28 Tage (oder 5 Halbwertszeiten, je nachdem, was länger ist) vor der ersten Dosis der Studienbehandlung.</li> <li>• Behandlung für DT begonnen, nachdem die dokumentierte Progression der DT festgestellt wurde (Einschlusskriterium „histologisch bestätigte DT mit einer Progression von <math>\geq 20 \%</math> innerhalb von 12 Monaten vor der Screening-Visite“).</li> <li>• MRT nicht tolerierbar oder kontraindiziert.</li> </ul>
<b>Intervention und Zahl der Patientinnen und Patienten</b>	<p>Gescreent gesamt: N = 201                      Randomisiert/Eingeschlossen: N = 142                      Intervention (Nirogacestat): N = 70                      Kontrolle (Placebo): N = 72</p>

Charakteristikum	Beschreibung
<b>Ort und Zeitraum der Durchführung; Datenschnitte</b>	<p><b>Studienzentren</b> 52 Studienzentren in 7 Ländern (Belgien, Kanada, Deutschland, Italien, Niederlanden, Vereinigtes Königreich, USA).</p> <p><b>Studienzeitraum</b></p> <ul style="list-style-type: none"> <li>• Erste Person erste Randomisierung: 15.05.2019</li> <li>• Letzte Person letzte Randomisierung: 03.08.2020 (Angabe aus Modul 4)</li> </ul> <p><b>Datenschnitte</b> 07.04.2022 (präspezifiziert)<sup>2)</sup>. Die finale Datenbanksperre erfolgte am 19.12.2024 (Präspezifizierung unklar, keine Daten eingereicht).</p>
<b>Endpunkte gemäß Studienprotokoll</b>	<p><b>Primärer Endpunkt</b> PFS</p> <p><b>Sekundäre Endpunkte</b></p> <ul style="list-style-type: none"> <li>• Sicherheit und Verträglichkeit</li> <li>• ORR (CR und PR)</li> <li>• Dauer des Ansprechens</li> <li>• Veränderung des Tumolvolumens</li> <li>• PRO (GODDESS, BPI-SF, PROMIS PF, EORTC QLQ-C30)</li> </ul> <p><b>Explorative Endpunkte</b></p> <ul style="list-style-type: none"> <li>• PGIS und PGIC</li> <li>• Häufigkeit und Verteilung von Mutation</li> <li>• Expression vor und nach der Dosis bei Notch-Signalweg-Genen</li> <li>• MRT-T2-Intensität</li> <li>• Beziehung zwischen Exposition und primären, sekundären und/oder explorativen Wirksamkeits- und Sicherheitsendpunkten</li> <li>• Inzidenz und Häufigkeit von Ereignissen, die möglicherweise einen Krankenhausaufenthalt aufgrund eines Dünndarmverschlusses oder aufgrund von DT-bedingten Schmerzen oder einer Operation für DT beinhalteten</li> </ul>

<sup>1)</sup> Für Personen, die sowohl eine intra- als auch eine extra-abdominale Zieltumorlokalisierung aufweisen, werden zu den intra-abdominalen Zieltumorlokalisierung klassifiziert.

<sup>2)</sup> Abschluss der primären Analyse nach 49 PFS-Ereignissen (ursprünglich 51 PFS-Ereignisse geplant), da zwischen 20.09.2021 und 07.04.2022 keine neuen Ereignisse auftraten.

<sup>3)</sup> Person war in die DB-Phase eingeschlossen und die primäre PFS-Analyse ist abgeschlossen oder Person im Placebo-Arm hat gemäß zentraler Bildauswertung eine radiographisch bestätigte fortschreitende Erkrankung (gemäß RECIST v1.1) oder Person im Nirogacestat-Arm hat gemäß zentraler Bildauswertung eine radiographisch bestätigte fortschreitende Erkrankung (gemäß RECIST v1.1), aber Person hat weiterhin klinischen Nutzen ohne signifikante Toxizität (Einschätzung durch Prüfpersonal) oder adäquate Organ- und Knochenmarkfunktion, wie in den Einschlusskriterien der RCT-Phase beschrieben (basierend auf Hämatologie- und Serumchemie-Ergebnissen innerhalb von 14 Tagen vor der Einschreibung in die OLE-Phase).

Abkürzungen: ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; BPI-SF: Brief Pain Inventory – Short Form; CR: Vollständiges Ansprechen; DT: Desmoidtumor; DB: doppelblind; ECOG: Eastern Cooperative Oncology Group; eCRF: electronic Case Report Form; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; EOT: End of Treatment; GODDESS: GOunder/DTRF DESmoid Symptom/Impact Scale; MRT: Magnetresonanztomographie; NSAID: Nichtsteroidale Antirheumatika; OLE: Open-Label-Extension; NYHA: New York Heart Association; ORR: Gesamtansprechrate; PFS: Progressionsfreies Überleben; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PR: Partielles Ansprechen; PRO: Patientenberichteter Endpunkt; PROMIS PF: Patient-Reported Outcomes Measurement Information System Physical Function; RCT: Randomized Controlled Trial; RECIST: Response Evaluation Criteria In Solid Tumors; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis; ULN: Upper Limit of Normal.

## Protokolländerungen

Das ursprüngliche globale Studienprotokoll wurde am 03.08.2018 abgeschlossen und fünfmal global geändert. Die globale Protokolländerung 4 wurde der „U.S. Food and Drug Administration“ (FDA) vorgelegt, aber nicht umgesetzt. Hinzu kommen länderspezifische Änderungen für u. a. Belgien, Italien und Deutschland. Wie viele Personen zum Zeitpunkt der Protokolländerungen tatsächlich eingeschlossen und damit unter Behandlung waren, konnte in den Studienunterlagen nicht identifiziert werden. Tabelle 3 ist zu entnehmen, wie viele Personen zu den jeweiligen Protokolländerungen in die Studie einwilligten. Dies waren unter Protokollversion 1 (27.11.2018) 113 Personen. Die erste Person wurde am 15.05.2019 in die Studie randomisiert.

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

<b>Amendment</b>	<b>Wesentliche Änderungen</b>
Version 2 vom 14.10.2019 (Anzahl Personen, die einwilligten: 25)	Die Stichprobengröße wurde von 105 auf 135 zu screenende Personen und von 94 auf 118 zu randomisierende Personen erhöht.
Version 3 vom 27.01.2020 (Anzahl Personen, die einwilligten: 63)	Als UESI wurden <ul style="list-style-type: none"> <li>• Störungen des Reproduktionssystems hinzugefügt, einschließlich Amenorrhö und vorzeitige Menopause / primäre Ovarialinsuffizienz, um eine zusätzliche Sicherheitsnachverfolgung zu ermöglichen;</li> <li>• gastrointestinale Ereignisse wie Übelkeit, Erbrechen/Dyspepsie und Durchfall entfernt, da diese als bekannte und erwartete Ereignisse im Zusammenhang mit Nirogacestat eingestuft wurden.</li> </ul>
Version 5 vom 09.02.2021 (Anzahl Personen, die einwilligten: 0)	Definition von PFS wurde überarbeitet, um Ereignisse klinischer Progression in die Analyse einzubeziehen.

Abkürzungen PFS: Progressionsfreies Überleben; UESI: Unerwünschtes Ereignis von besonderem Interesse.

## Charakterisierung der Intervention

Tabelle 4: Charakterisierung der Intervention in der Studie DeFi

Intervention	Kontrolle
<p><b>Dosierung</b> Nirogacestat 150 mg (3 x 50 mg Tabletten) zweimal täglich je Zyklus (28 Tage).</p> <p><b>Dosisanpassungen von 150 mg auf 100 mg zweimal täglich bei</b></p> <ul style="list-style-type: none"> <li>• anhaltender Diarrhö, anhaltender Übelkeit oder anhaltendem Erbrechen jeweils <math>\geq</math> Grad 3 für <math>\geq</math> 3 Tage trotz maximaler medikamentöser Therapie</li> <li>• vorzeitiger Menopause / primärer Ovarialinsuffizienz <math>\geq</math> Grad 2a<sup>1)</sup></li> <li>• Toxizität der Haut <math>\geq</math> Grad 3<sup>2)</sup></li> <li>• Hypophosphatämie <math>\geq</math> Grad 3, die trotz maximaler Ersatztherapie und in Abwesenheit von Symptomen <math>\geq</math> 7 Tage anhält</li> <li>• klinisch signifikanten nicht-hämatologischen Toxizitäten <math>\geq</math> Grad 3</li> <li>• hämatologischen Toxizitäten <math>\geq</math> Grad 3</li> </ul> <p><b>Regeln zum Abbruch der Studienmedikation bei</b></p> <ul style="list-style-type: none"> <li>• Anaphylaxie</li> <li>• Hypersensitivitätsreaktion <math>\geq</math> Grad 3</li> <li>• Leber-Toxizitäten</li> </ul> <p>Die Studienbehandlung wurde erst nach Abklingen des UE auf <math>\leq</math> Grad 1 oder Baseline fortgesetzt. Bei Abklingen innerhalb von 14 Tagen erfolgte die Wiederaufnahme mit reduzierter Dosis. Konnte das UE nach 22 Tagen nicht ausreichend abklingen, war eine Rücksprache mit dem Sponsor erforderlich, um die Behandlung wiederaufzunehmen. Ein dauerhafter Abbruch erfolgte nach Rücksprache, wenn dasselbe UE <math>\geq</math> Grad 3 unter reduzierter Dosis erneut und therapiebedingt auftrat.</p>	<p><b>Dosierung</b> Placebo 150 mg (3 x 50 mg Tabletten) zweimal täglich je Zyklus (28 Tage).</p>
<p><b>Nicht erlaubte Begleitmedikation</b> Tägliche Einnahme von NSAID zur Behandlung von DT<sup>3)</sup>, Tyrosinkinase-Hemmer, andere antineoplastische Therapien, einschl. zytotoxischer Substanzen, zielgerichteter Agenzien, endokriner Therapie oder anderer Antikörper; sowie jede experimentelle Behandlung für DT.</p> <p><b>Erlaubte Begleitmedikation</b></p> <ul style="list-style-type: none"> <li>• Phosphat-Supplemente bei Nirogacestat-assoziiertes Hypophosphatämie; Art und Dauer der Behandlung lag im Ermessen des Prüfpersonals.</li> <li>• Medikamente zur Standardbehandlung der Symptome oder zur supportiven Behandlung der Auswirkungen der Studienbehandlung konnten nach Ermessen des Prüfpersonals verabreicht werden, es sei denn, sie waren als Begleitmedikationen ausgeschlossen.</li> <li>• Die gelegentliche Verwendung von NSAID (<math>\leq</math> 3 Tage pro Woche) für schmerz- oder entzündungshemmende Zwecke war während der gesamten Behandlungsdauer erlaubt.</li> <li>• Anti-emetische und anti-diarrhöische Medikamente konnten nach Ermessen des Prüfpersonals verwendet werden, vorausgesetzt es gab keine bekannten oder erwarteten Arzneimittelwechselwirkungen.</li> </ul>	

<sup>1)</sup> Eine Dosisreduktion ist bei Ereignissen von vorzeitiger Menopause / primärer Ovarialinsuffizienz nicht erforderlich, kann jedoch für symptomatische Teilnehmende auf Grundlage des individuellen Nutzen-Risiko-Profiles in Betracht gezogen werden.

<sup>2)</sup> Gemäß Fachinformation umfassen Hautreaktionen: Follikulitis, makulopapulöser Ausschlag und Hidradenitis [17].

<sup>3)</sup> Personen, die chronisch NSAID zur Behandlung anderer Erkrankungen als DT einnehmen, müssen diese bereits vor der dokumentierten fortschreitenden DT-Erkrankung erhalten haben und müssen die Dosis 28 Tage vor der ersten Gabe der Studienmedikation stabil beibehalten haben. Die gelegentliche Einnahme (definiert als  $\leq$  3 Tage pro Woche) von NSAID zur Behandlung von Schmerzen oder als Entzündungshemmer bei Erkrankungen wie Kopfschmerzen, Arthritis usw. ist während der gesamten Behandlungsphase der Studie erlaubt. Dosissteigerungen von NSAID sind während der Behandlungsphase in der DB-Phase der Studie nicht gestattet.

Abkürzungen: DB: doppelblind; DT: Desmoidtumor; NSAID: Nichtsteroidale Antirheumatika; 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 DeFi*

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Todesfälle <sup>1)</sup>	Mortalität	Ja	Ja
PFS <sup>2)</sup>	Morbidity	Ja	Ergänzend
<u>Ansprechen</u> ORR Veränderung des Tumorzvolumens		Ja	Nein
GODDESS-DTSS/-DTIS <sup>3)</sup>		Ja	Ja
BPI-SF		Ja	Ja
PROMIS PF Short Form 10a/13		Ja	Nein
PGIS		Ja	Nein <sup>4)</sup>
PGIC		Ja	Ja
EORTC QLQ-C30: Symptomskalen		Ja	Ja
EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus	Lebensqualität	Ja	Ja
Unerwünschte Ereignisse	Sicherheit	Ja	Ja

<sup>1)</sup> Todesfälle wurden unter Sicherheit erfasst.

<sup>2)</sup> Primärer Endpunkt (Jegliche qualifizierte klinische Progression oder Tod).

<sup>3)</sup> Die DTIS wird abweichend von der Einschätzung des pU der Endpunktkategorie „Morbidity“ zugeordnet.

<sup>4)</sup> Die vorgelegte Auswertung in Modul 4 zu Zyklus 24 mittels MMRM wird aufgrund der geringen Rücklaufquoten (bereits zu Zyklus 4 < 70%) als nicht adäquat eingeschätzt.

Abkürzungen: Brief Pain Inventory – Short Form; DTIS: Desmoid Tumor Impact Scale; DTSS: Desmoid Tumor Symptom Scale; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; MMRM: Mixed Model for Repeated Measures; ORR: Gesamtansprechrate; PFS: Progressionsfreies Überleben; 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*

Die während der Studie aufgetretenen Todesfälle jeglicher Ursache wurden im Rahmen der Sicherheitserfassung kontinuierlich erfasst (siehe Kapitel 2.3.4). Gemäß Modul 4 wurde das Gesamtüberleben definiert als Zeit (in Monaten) vom Zeitpunkt der Randomisierung bis zum Datum des Todes. Wenn der Tod innerhalb von 30 Tagen nach der letzten Studienbehandlung oder vor Abbruch der Studienbehandlung (je nachdem, was zuletzt eintrat) eintrat, erfolgte die Erfassung des Todes.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Eine Erfassung der Todesfälle über den Progress hinaus erfolgte nur für 30 Tage für Personen, die nicht in die Open-Label-Extension (OLE)-Phase übergangen oder innerhalb der OLE.

#### Patientenrelevanz

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

#### Validität

Die Erhebung des Endpunkts wird als valide erachtet, sofern sich die Personen unter Behandlung befanden.

### 2.3.2 Morbidität

#### PFS

Der Endpunkt „PFS“ wird in der Operationalisierung des „klinischen progressionsfreien Überlebens“ aufgrund von Unklarheiten in der Operationalisierung und damit einhergehender unklarer Patientenrelevanz ergänzend in der Nutzenbewertung dargestellt.

#### Operationalisierung

##### *Beschreibung*

„Progressionsfreies Überleben“ (PFS) wurde als primärer Endpunkt in der Studie DeFi präspezifiziert als die Zeit (in Monaten) zwischen dem Zeitpunkt der Randomisierung bis zum Datum der Feststellung einer radiographischen Progression oder Tod jeglicher Ursache (je nachdem, was zuerst eintrat). Die Beurteilung der radiographischen Progression erfolgt objektiv mittels bildgebender Verfahren (Magnetresonanztomographie (MRT) oder Computertomographie) gemäß „Response Evaluation Criteria In Solid Tumors“ (RECIST; v1.1) [1]. Für Progressionsereignisse ist das Datum der Bildgebung maßgeblich. Die Ereignisse wurden basierend auf der Bildgebung durch das zentrale Bildgebungszentrum qualifiziert.

Mit Protokollamendment 5 vom 09.02.2021 wurde zudem der Aspekt der klinischen Progression in die Definition aufgenommen. Damit konnte die Feststellung einer Progression radiographisch oder klinisch erfolgen. Als Kriterien für die Feststellung einer klinischen Progression galten das Auftreten oder die Verschlechterung von Symptomen, die zu einer globalen Verschlechterung des Gesundheitszustands führten und damit den Abbruch der Studienbehandlung sowie den Beginn einer anschließenden Therapie zur Folge hatten. Das Datum der klinischen Progression ist dabei das früheste Datum des Auftretens oder der Verschlechterung von Symptomen, welche

zu einer globalen Verschlechterung des Gesundheitszustands führten. Die klinische Progression wurde durch ein unabhängiges und verblindetes „Endpoint Adjudication Committee“ (EAC) dahingehend geprüft, ob sich die Ereignisse als klinische Progression hinsichtlich des Einschlusses in den Endpunkt „PFS“ gemäß einer EAC-Überprüfungsrichtlinie qualifizieren. UE und SUE, die mit einer klinischen Progression assoziiert sind, sowie Medikationen und Prozeduren, die für die Behandlung von Desmoidtumoren innerhalb von 30 Tagen nach der letzten Dosis der Studienmedikation auftraten, wurden zudem im „electronic Case Report Form“ dokumentiert.

Zu folgenden Zeitpunkten erfolgte eine Zensurierung von Patienten:

- Datum der Randomisierung, wenn keine adäquate Bewertung des Krankheitsstatus vorlag (letzte Bildgebung).
- Bei fehlender Dokumentation von Progression oder Tod: Datum der letzten adäquaten Bewertung des Krankheitsstatus.
- Frühzeitiger Studienabbruch durch Prüfpersonal aufgrund einer klinischen Progression, welche nicht die Definition eines qualifizierten Ereignisses erfüllt (Beurteilung durch EAC): Datum der letzten adäquaten Bewertung des Krankheitsstatus.
- Start einer neuen Antikrebstherapie oder eines neuen Verfahrens vor der dokumentierten radiographischen oder klinischen Progression: Datum der letzten adäquaten Bewertung des Krankheitsstatus.

### *Bewertung*

Die Kriterien zur Operationalisierung einer radiographisch bestimmten Progression werden mittels RECIST klar definiert. Die Operationalisierung hinsichtlich der klinischen Progression ist hingegen unklar. In Modul 4 führt der pU auf, dass die klinische Progression den Beginn einer anschließenden Behandlung erforderlich macht. Als mögliche Behandlungsoptionen werden u. a. Strahlentherapie, chirurgischer Eingriff oder systemische Therapien wie Chemotherapie oder Tyrosinkinase-Inhibitoren genannt. Konkrete Kriterien, anhand derer das Prüfpersonal diese klinische Progression definitiv feststellen soll, werden jedoch nicht näher spezifiziert. Dadurch bleibt bspw. unklar, anhand welcher Kriterien eine globale Verschlechterung des Gesundheitszustands bestimmt wurde und inwieweit eine systematische Erhebung erfolgte. Klinische Progressionsereignisse wurden zwar anhand einer Überprüfungsrichtlinie durch ein EAC qualifiziert, allerdings sind auch in diesem Kontext die Kriterien nicht näher beschrieben und die Richtlinie wurde nicht mit dem Dossier vorgelegt. Zudem ist unklar, inwieweit alle Personen hinsichtlich einer klinischen Progression bewertet werden konnten, da der Endpunkt erst mit Protokollamendment 5 ca. 21 Monate nach Einschluss der ersten Person eingeführt wurde.

Die Nicht-Berücksichtigung von Tumorbewertungen, die nach dem Beginn einer nachfolgenden Anti-Tumortherapie durchgeführt wurden, entspricht nicht dem Treatment-Policy-Ansatz.

### Patientenrelevanz

Der Endpunkt „PFS“ ist ein kombinierter Endpunkt aus Überleben (Mortalität) und Krankheitsprogression (Morbidität). Die Komponente „Überleben“ ist ein patientenrelevanter Teilaspekt und wird bereits im Endpunkt „Todesfälle“ in der Nutzenbewertung berücksichtigt.

Die Morbiditätskomponente „Krankheitsprogression“ wird unterschiedlich operationalisiert. Zum einen basiert die Feststellung einer Progression auf der Beurteilung radiographischer Befunde mittels RECIST durch medizinisches Fachpersonal. Hierbei besteht kein Bezug zur Symptomatik der betreffenden Personen, wodurch keine unmittelbare Patientenrelevanz abgeleitet werden kann. Für die klinische Progression hingegen galt das Auftreten oder die Verschlechterung von Symptomen, die zu einer globalen Verschlechterung des Gesundheitszustands führten und das Ende der Studienbehandlung und den Beginn einer Notfallbehandlung zur Folge hatten, als Progress. Vor dem Hintergrund, dass die Erfassung des Symptombezugs unklar ist und die Notwendigkeit einer

Notfalltherapie nicht klar als patientenrelevant zu werten ist, wird die klinische Progression in ihrer Patientenrelevanz als unklar bewertet.

### Validität

Aufgrund von umfangreichen Unklarheiten in der Operationalisierung kann die Validität nicht abschließend bewertet werden. Eine fehlende standardisierte Erhebung kann die Validität einschränken.

### **Tumoransprechen**

Die Endpunkte zum Tumoransprechen – „Gesamtansprechrte“, „Veränderung des Tumorumens“ – werden in der Nutzenbewertung aufgrund nicht-gegebener Patientenrelevanz nicht berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Die Gesamtansprechrte (ORR) wurde analog der radiographischen Progression durch zentrale Gutachter gemäß RECIST beurteilt [1]. ORR wurde dabei als Anteil der Personen mit einem vollständigen Ansprechen (CR) und einem partiellen Ansprechen (PR) definiert. PR und CR mussten durch einen nachfolgenden Scan bestätigt werden. Dieser erfolgte frühestens 4 Wochen nach der ersten Feststellung einer PR oder CR. In diesem Kontext wurde CR als Verschwinden aller Zielläsionen definiert und PR als eine mindestens 30%ige Verringerung der Summe der längsten Durchmesser der Zielläsionen, wobei die Summe der längsten Durchmesser zu Baseline die Referenzgröße darstellte.

Die Veränderung des Tumorumens war ursprünglich als sekundärer Endpunkt präspezifiziert und wurde im Verlauf der Studie zu einem explorativen Endpunkt geändert. Die Bewertung der Veränderung des größten Zieltumors erfolgte anhand volumetrischer MRT-Scans. Die Referenz stellte der größte Zieltumor zu Baseline dar. Es erfolgte eine Übermittlung aller Scans an das zentrale Bildgebungszentrum, in welchem eine Bewertung durch die zentrale Bildgebungsüberprüfung durchgeführt wurde. Für diesen Endpunkt wurde keine Responderschwelle präspezifiziert.

#### *Bewertung*

Die Operationalisierung der Endpunkte ist nachvollziehbar.

### Patientenrelevanz

Bezogen auf die Gesamtansprechrte führt der pU in Modul 4 aus, dass das Ansprechen auf die Therapie oft mit einer Symptomlinderung einhergeht. Die hier vorliegende Operationalisierung basiert jedoch ausschließlich auf bildgebenden Verfahren. Eine Veränderung einer für die Patientin / den Patienten spürbaren Symptomatik findet im Endpunkt keine direkte Berücksichtigung. Daher wird der Endpunkt als nicht unmittelbar patientenrelevant eingeschätzt.

Die Veränderung des Tumorumens verbunden mit einer für die Patientin / den Patienten spürbaren Abnahme von Krankheitssymptomen ist für die Nutzenbewertung patientenrelevant. Der pU stellt in Modul 4 des Dossiers zu Nirogacestat dar, dass Desmoidtumore zu einer Kompression von Vitalstrukturen führen können, was unter anderem Schmerzen, Funktionseinschränkungen und weitere Komplikationen zur Folge haben kann. Eine Reduktion des Tumorumens könne daher positive Auswirkungen auf die Symptomatik haben. Zudem könnte die Veränderung des Tumorumens die Lebensqualität bei äußerlich sichtbaren Tumoren beeinflussen. Jedoch ist in der Studie DeFi anhand der Ein- und Ausschlusskriterien nicht sichergestellt, dass in allen Teilnehmenden zu Baseline eine entsprechende Symptomatik bzw. Funktionseinschränkung vorlag, sodass bei einer Reduktion des Tumorumens nicht zwangsläufig von einer Verbesserung der Symptomatik ausgegangen werden kann.

Die Erhebung des Endpunkts „Veränderung des Tumorzvolumens“ operationalisiert mittels bildgebender Verfahren wird nicht als unmittelbar patientenrelevant angesehen. Die Erhebung des Endpunkts erfolgt nicht in Zusammenhang mit einer für die Patientin / den Patienten spürbaren Symptomatik.

### Validität

Aufgrund der fehlenden Patientenrelevanz wird die Validität der Endpunkte nicht beurteilt.

### **Symptomatik mittels GODDESS-DTSS/-DTIS**

Die Endpunkte „GODDESS-DTSS“ und „GODDESS-DTIS“ werden in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Die „GOunder/DTRF DEsmoid Symptom/Impact Scale“ (GODDESS) ist ein patientenberichtetes Instrument mit 28 Items. Das Instrument wurde krankheitsspezifisch zur Quantifizierung von Symptomen von Desmoidtumoren sowie deren Auswirkungen auf Patientinnen und Patienten entwickelt. Die Erfassung der GODDESS erfolgt, wie bei allen PRO, zuhause über elektronische PRO-Geräte.

Die GODDESS gliedert sich in 2 Subskalen, die „Desmoid Tumor Symptom Scale“ (DTSS) und die „Desmoid Tumor Impact Scale“ (DTIS). Es wird kein skalenübergreifender Gesamtscore gebildet. Items, welche keiner Domäne zugeordnet wurden, werden als Einzelscores betrachtet.

Die Subskala DTSS wird der Endpunktkategorie „Morbidität“ zugeordnet und besteht aus 11 Items, die der Beurteilung der Desmoidtumor-bedingten Schlüsselsymptome sowie deren Schweregrad dienen. Hierzu gehören: Schmerz, Fatigue, Schwellung, Muskelschwäche, Bewegungsschwierigkeiten, Übelkeit und Völlegefühl.

Die Items werden auf einer 11-stufigen numerischen Bewertungsskala (NRS) von 0 („nicht vorhanden“) bis 10 („so schlimm, wie nur vorstellbar“) bewertet. Höhere Werte implizieren eine stärkere Symptomschwere.

Die DTSS beinhaltet 3 Domänen („Schmerz“ (Items 1–3), „Extra-abdominale Symptome“ (Items 5–7), „Intra-abdominale Symptome“ (Items 9–11)) sowie die Einzelitems „Fatigue“ und „Tumorklassifikation“. Die Abfrage der Domäne „Intra-abdominale Symptome“ erfolgt nur, wenn die entsprechende Tumorklassifikation in Item 8 angegeben wurde.

Zur Auswertung wurden für die DTSS tägliche Gesamtscores und wöchentliche Durchschnittsscores berechnet. Insgesamt werden 6 verschiedene Scores zur DTSS berechnet, wobei kein DTSS-Gesamtscore basierend auf allen Items berechnet wird:

- Gesamtscore der Symptome („Gewichtete DTSS-Gesamtscore“) – 7 Items
- Gesamtscore der Symptome – 5 Items (ungewichtet)
- Gesamtscore der Symptome („Ungewichtete DTSS-Gesamtscore“) – 7 Items
- Domänenscore „Schmerz“ – 3 Items
- Domänenscore „Extra-abdominale Symptome“ – 3 Items
- Domänenscore „Intra-abdominale Symptome“ – 3 Items

Die Recall-Periode der DTSS-Items beträgt 24 Stunden, der DTSS wird in der Woche vor dem nächsten Zyklus bzw. vor der Follow-up-Visite täglich erhoben. Im Falle fehlender Werte auf Item-Ebene wird kein Score berechnet. Die täglichen Scores wurden über eine Woche gemittelt, wenn tägliche Scores von mindestens 4 der 7 Tage vorlagen.

Die Subskala DTIS, bestehend aus 17 Items, quantifiziert den Einfluss der krankheitsspezifischen Symptomatik auf Funktionsfähigkeit und Aktivitäten des täglichen Lebens. Der pU ordnet den Endpunkt der Endpunktkategorie „Lebensqualität“ zu.

Die Items werden mittels einer 11-Punkte-NRS zur Einschätzung des Schweregrads oder mittels einer 5-Punkte-Likert-Skala von „zu keinem Zeitpunkt“ bis „die ganze Zeit“ zur Messung der Häufigkeit beantwortet. Die Recall-Periode beträgt 7 Tage. Höhere Werte bedeuten einen stärkeren Einfluss der Symptomatik auf Aktivitäten des täglichen Lebens.

Für die DTIS wird kein übergreifender Gesamtscore gebildet, stattdessen werden 3 Domänenscores gebildet:

- Domänenscore „Körperliche Funktion“ (5-Punkte-Likert-Skala) – 5 Items
- Domänenscore „Schlaf“ (5-Punkte-Likert-Skala) – 3 Items
- Domänenscore „Emotion“ (11-Punkte-NRS) – 6 Items
- Einzelscores zu Item 9, 10 und 11 (Fragen zu „Körperwahrnehmung“, „Selbstbewusstsein“ und „Alltägliche praktische Herausforderungen“)

### *Bewertung*

Die Operationalisierung der GODDESS-DTSS ist nachvollziehbar. Der pU reicht mit dem Dossier Auswertungen zu insgesamt 3 verschiedenen Gesamtscores der Symptome ein, welche ähnliche Items mit teilweise unterschiedlicher Gewichtung abbilden. Aufgrund der gewichteten Berücksichtigung der Anzahl an Items zur Schmerz-Symptomatik wird in der vorliegenden Nutzenbewertung der „Gewichtete DTSS-Gesamtscore“ berücksichtigt. Dieser erscheint adäquat, um eine übermäßige Gewichtung der Schmerzdimension zu vermeiden und wird in den eingereichten Unterlagen zur Validität empfohlen. Daneben wird der Domänenscore „Intra-abdominale Symptome“ berücksichtigt.

Die Operationalisierung der GODDESS-DTIS ist nachvollziehbar beschrieben. Die ausgewerteten DTIS-Domänenscores umfassen Items, die der physischen (Domänenscore „Körperliche Funktion“ und „Schlaf“) und psychischen Dimension (Domänenscore „Emotion“) zuzuordnen sind. Die soziale Dimension wird in der DTIS lediglich durch ein Einzelitem adressiert, welches vom pU jedoch nicht ausgewertet wurde. Dies wird als nicht adäquat bewertet. Aufgrund der fehlenden Berücksichtigung der sozialen Dimension, welche integraler Bestandteil des multidimensionalen Konstrukts der Lebensqualität ist, wird der Endpunkt abweichend von der Einschätzung des pU der Endpunktkategorie „Morbidity“ zugeordnet.

### Patientenrelevanz

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

### Validität

Das Erhebungsinstrument wurde prospektiv an 31 Personen mit Desmoidtumoren unter Nutzung adäquater qualitativer Methoden entwickelt [6]. Basierend auf den vorliegenden Informationen kann insgesamt von einer hinreichenden Inhaltsvalidität des Erhebungsinstruments ausgegangen werden.

Zur Untersuchung der psychometrischen Eigenschaften des Fragebogens, insbesondere Validität und Reliabilität, wurden gepoolte Analysen beider Studienarme der Studie DeFi genutzt, welche der vorliegenden Nutzenbewertung zugrunde liegt [5]. Die fehlende Validierung in einer separaten Studie und die dadurch fehlende abgeschlossene Entwicklung des Instruments vor dem Einsatz in der Studie DeFi ist als Limitation anzusehen.

Die GODDESS-Domänenstruktur konnte durch Korrelationsanalysen und durch eine konfirmatorische Faktoranalyse bestätigt werden. Die interne Konsistenz, quantifiziert mittels Cronbach's alpha, wird als überwiegend gut bis exzellent bewertet, wobei die Werte in der Domäne „Intra-abdominale Symptome“ der DTSS geringer waren. Für den gewichteten DTSS-Gesamtscore und die Domäne „Schmerz“ der DTSS lagen die Werte für Cronbach's alpha bei 0,95 oder höher, was einen Hinweis

auf möglicherweise redundante Items liefert. Die Test-Retest-Reliabilität für den gewichteten DTSS-Gesamtscore sowie die Subdomänen wird als exzellent eingeschätzt ( $ICC > 0,90$ ) [5].

Die Konstrukt- und Known-Groups-Validität wurde durch adäquate Analysen unterstützt. Die Änderungssensitivität konnte hinsichtlich von Verbesserungen gezeigt werden. Die Anzahl von Personen mit Verschlechterungen ( $n = 5-10$ ) war zu gering, um zuverlässige Aussagen treffen zu können. Für Items der DTSS wurden zu Baseline häufig Bodeneffekte ( $> 20\%$  der Studienpopulation berichteten keine Symptomatik für 7 der 10 Items) beobachtet, was ebenfalls auf mögliche Redundanzen hinweisen könnte und die Beobachtung von Verbesserungen in der Symptomatik einschränkt [5].

Die Struktur der DTIS-Domänen wird grundsätzlich durch die Inter-Item-Korrelationsanalysen unterstützt. Dabei ist in der Domäne „Körperliche Funktion“ auf sehr hohe Korrelationen ( $r = 0,93$ ) zwischen dem Item „Heranreichen (Häufigkeit)“ und dem Item „Heranreichen (Schwierigkeiten)“ hinzuweisen, was einen Hinweis auf mögliche Redundanzen liefert. Laut Angaben in der Validierungsstudie korrelierten die Einzelscores 9 und 11 nicht hoch mit anderen Items, weswegen diese keiner spezifischen Domäne zugeordnet wurden. Zusammenfassend wurden die Items 9, 10 und 11 keiner Domäne zugeordnet und keiner weiteren psychometrischen Beurteilung unterzogen, was kritisch bewertet wird, da diese Items bei der Entwicklung als relevante Items der DTIS identifiziert wurden [5].

Die Konstrukt- und Known-Groups-Validität wurde nachgewiesen. Die interne Konsistenz lag für die DTIS-Subdomänen in einem guten bis exzellenten Bereich (Cronbach's alpha:  $0,87-0,95$ ). Insbesondere die Werte für die Domäne „Emotion“ lagen in einem sehr hohen Bereich (Cronbach's alpha:  $0,92-0,95$ ), was gegebenenfalls auf Redundanzen in dieser Domäne hinweisen könnte. Die Test-Retest-Reliabilität wurde zwischen Baseline und Zyklus 2 (Woche 8) bei den Personen berechnet, deren PGIS-Wert unverändert blieb. Es wurden gute Übereinstimmungen berichtet ( $ICC > 0,85$ ). Für die DTIS konnten sowohl Bodeneffekte (für 10 der 17 Items berichteten  $> 20\%$  der Studienpopulation keine Einschränkung) als auch Deckeneffekte (für 2 der 17 Items (Items 5 und 6) berichteten  $> 20\%$  der Studienpopulation die maximale Einschränkung) beobachtet werden. Die Domänen des DTIS zeigten eine Änderungssensitivität hinsichtlich einer Verbesserung. Die Anzahl von Personen mit Verschlechterungen ( $n = 10$ ) war jedoch zu gering, um zuverlässige Aussagen treffen zu können [5].

Im Rahmen der Nutzenbewertung wird die im Dossier verwendete Responderschwelle von  $15\%$  der Skalenspannweite (entspricht 1,65 Punkten) als adäquat eingeschätzt.

### **Schmerz mittels BPI-SF**

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

#### Operationalisierung

##### *Beschreibung*

Der „Brief Pain Inventory – Short Form“ (BPI-SF) wurde in der Studie DeFi als sekundärer Endpunkt erfasst und dient der Bewertung von klinischen Schmerzsymptomen. Der BPI-SF besteht aus insgesamt 9 Items, welche eine Beurteilung des Schweregrads der Schmerzen sowie des Ausmaßes, in welchem die Schmerzen Aktivitäten des täglichen Lebens beeinflussen, ermöglichen.

Die Beantwortung der Items erfolgt mittels einer 11-Punkte-NRS von 0 bis 10 mit einer 24-stündigen Recall-Periode. Für die Items 3–6 (Schmerzintensität) bedeutet 0 = „keine Schmerzen“ und 10 = „stärkste vorstellbare Schmerzen“, für die Items 9A–9G (Beeinträchtigung durch den Schmerz) 0 = „keine Beeinträchtigung“ und 10 = „vollständige Beeinträchtigung“.

Zur Messung der durchschnittlichen Schmerzintensität werden ein täglicher sowie ein wöchentlicher Score berechnet („Pain Severity Subscale“). Tägliche Scores bestehen aus dem Durchschnitt von 4 Antworten am Tag, der wöchentliche Score wird als Durchschnitt der täglichen Scores berechnet.

Für die Berechnung des täglichen Scores müssen alle 4 Items beantwortet werden, für den wöchentlichen Score müssen Werte aus mindestens 4 von 7 Tagen vorliegen. Zur Messung der durchschnittlichen Beeinträchtigung durch den Schmerz wird eine „Pain Interference Subscale“ gebildet. Auch hier wird zwischen täglichem und wöchentlichem Score unterschieden: Tägliche Scores werden als Durchschnitt aus den 7 Items gebildet, insofern 4 von 7 Items keine fehlenden Werte aufweisen. Der wöchentliche Score wird als Durchschnitt der täglichen Scores gebildet, insofern für 4 von 7 Tagen keine fehlenden Werte vermerkt sind.

Neben der „Pain Severity Subscale“ und der „Pain Interference Subscale“ legt der pU Daten für eine „Durchschnittliche Schmerzintensität“ (API) vor; dies ist der Durchschnitt des täglich gemessenen Item 3 des BPI-SF („Schlimmste Schmerzen in den letzten 24 Stunden“) über einen Zeitraum von 7 Tagen vor jeder Visite und wird nur gebildet, wenn für 4 von 7 Tagen keine fehlenden Werte vorliegen.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Auswertungen basierend auf dem Einzelitem 3 werden nicht dargestellt, da dieses bereits in der „Pain Severity Subscale“ enthalten ist und diese die Schmerz-Symptomatik umfassender abbildet.

#### Patientenrelevanz

Der Endpunkt „BPI-SF“ wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen

#### Validität

Die Validität des Endpunkts „BPI-SF“ wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet und wird im zugrundeliegenden Anwendungsgebiet als valide eingeschätzt. In der Nutzenbewertung werden die eingereichten Responderanalysen mit einer Responder Schwelle von  $\geq 15\%$  der Skalenspannweite berücksichtigt.

#### **Körperliche Funktion mittels PROMIS PF Short Form 10a/13**

Der Endpunkt „PROMIS PF Short Form 10a/13“ wird in der Nutzenbewertung aufgrund unzureichender Angaben zur Validität und Reliabilität sowie Doppelerfassung zu anderen Endpunkten nicht in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Zur Messung der körperlichen Funktionsfähigkeit wurde der „Patient-Reported Outcomes Measurement Information System Physical Function Short Form 10a“ (PROMIS PF Short Form 10a), Version 2.0, verwendet. In der Studie DeFi wurde das Instrument als explorativer Endpunkt erhoben. Wie Modul 4 zu entnehmen ist, wurde der Endpunkt aufgrund von Doppelerfassung mit anderen PRO von einem sekundären in einen explorativen Endpunkt im Studienprotokoll/SAP geändert. Der PROMIS PF Short Form 10a besteht aus insgesamt 10 Items, die sich auf eine Recall-Periode von 7 Tagen beziehen. Der Fragebogen dient zur Messung der Fähigkeit, körperliche Aktivitäten auszuführen und umfasst die oberen und unteren Extremitäten, zentrale Körperregionen sowie Aktivitäten des täglichen Lebens. Die Items werden mittels einer 5-Punkte-Skala bewertet [„not at all – cannot do“ bzw. „without any difficulty – cannot do“].

Neben dem PROMIS PF Short Form 10a wurden zusätzlich 3 weitere Items zur körperlichen Funktion aus der PROMIS-Item-Datenbank ausgewählt:

- Zusatzitem 1: „Können Sie Ihren Rücken beugen oder drehen?“
- Zusatzitem 2: „Können Sie in einen hohen Schrank greifen?“
- Zusatzitem 3: „Ich habe Schwierigkeiten, meine regelmäßigen täglichen Arbeiten im Haus zu erledigen.“

Nach Angaben in Modul 4 erfolgt die Auswertung des PROMIS PF Short Form 10a als Gesamt-Summenscore. Darüber hinaus wird ein standardisierter T-Score und ein Standardfehler des T-Scores, bezugnehmend auf den Gesamtscore, berechnet. Dieser basiert auf einer Konvertierungstabelle. Ein höherer T-Score steht für eine bessere körperliche Funktionsfähigkeit, ein niedrigerer für eine eingeschränkte Funktionsfähigkeit im Vergleich zur Normpopulation. Zudem erfolgt die Auswertung eines „PROMIS PF 13“-Scores als Summenwert mit den 3 zusätzlichen Items, insofern keine fehlenden Werte zu einzelnen Items vorlagen.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar. Der Endpunkt wird abweichend vom pU der Endpunktkategorie „Morbidity“ zugeordnet, da er aus dem multidimensionalen Konstrukt der Lebensqualität nur die physischen Aspekte erfasst.

#### Patientenrelevanz

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

#### Validität

Der pU legt zur Untersuchung der psychometrischen Eigenschaften des PROMIS PF Short Form 10a einen psychometrischen Analysebericht vor, basierend auf den Daten der Studie DeFi [12]. Ferner verweist der pU auf eine Publikation zur PROMIS-Item-Datenbank im Kontext der körperlichen Funktionsfähigkeit [11]. Nach Angaben des pU erfolgte die Auswahl der 3 Zusatzitems durch Untersuchungen zur Inhaltsvalidität. Demnach basierte die Auswahl der Items darauf, welche weiteren Elemente der körperlichen Funktion als relevant eingeschätzt wurden. Eine genauere Beschreibung zu den angewandten Methoden sowie dem Einbezug von Patienten und Ergebnissen hierzu liegt nicht vor.

Die Informationen, die aus den eingereichten Unterlagen hervorgehen, sind nicht ausreichend, um die Validität und Reliabilität des PROMIS PF Short-Form 10a bzw. „PROMIS PF 13“ in der gegenständlichen Patientenpopulation zu bewerten. Darüber hinaus ist auf Überschneidungen mit weiteren Erhebungsinstrumenten der Studie DeFi hinzuweisen. Die körperliche Funktion wird bspw. bereits mit der krankheitsspezifischen GODDESS-DTIS sowie dem EORTC QLQ-C30 erfasst, wodurch, auch in Anbetracht der eingeschränkten psychometrischen Güte, kein zusätzlicher Informationsgewinn durch den PROMIS PF Short Form 10a bzw. „PROMIS PF 13“ entsteht.

#### **Desmoidtumor-bedingte Krankheitssymptome mittels PGIS**

Der Endpunkt „PGIS“ wird in der Nutzenbewertung nicht berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Der „Patient Global Impression of Severity“ (PGIS) wurde als explorativer Endpunkt erhoben. Es handelt sich um eine Einzelitem-Skala, welche den subjektiven Eindruck einer Person in Bezug auf den Gesamtschweregrad der Desmoidtumor-bedingten Krankheitssymptome erfasst. Die Skala bezieht sich auf die letzten 7 Tage (Recall-Periode). Die Beantwortung erfolgt auf einer 4-Punkte-Skala von „keine“ bis „stark“.

##### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

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

### Validität

Die Validität des Endpunkts „PGIS“ wurde in vorherigen Nutzenbewertungsverfahren bereits bewertet. Allerdings liegen keine verwertbaren Ergebnisse vor. Zudem ist nicht ersichtlich, warum als Bezugsauswertungspunkt Zyklus 24 herangezogen wurde. Responderanalysen, analog wie beim PGIC, wurden nicht eingereicht.

### **Allgemeiner Gesundheitszustand mittels PGIC**

Der Endpunkt „PGIC“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Der „Patient Global Impression of Change“ (PGIC) wurde als explorativer Endpunkt erhoben. Dabei handelt es sich um eine Einzelitem-Skala, welche eine Bewertung der Wahrnehmung der Patientinnen und Patienten hinsichtlich der Veränderung ihres allgemeinen Gesundheitsstatus im Vergleich zu Baseline ermöglicht. Der Bezugszeitraum stellt die Veränderung innerhalb der letzten 7 Tage im Vergleich zum Beginn der Studienbehandlung dar. Das Item wird auf einer 7-Punkte-Skala von „sehr viel besser“ bis „sehr viel schlechter“ bewertet.

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

### Patientenrelevanz

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

### Validität

Bei der rückwirkenden Einschätzung einer Veränderung müssen die Befragten in der Lage sein, sowohl ihren gegenwärtigen Zustand als auch ihren Ausgangszustand zu quantifizieren und daraus eine mentale Subtraktion durchzuführen. Dies scheint jedoch nicht möglich zu sein. Stattdessen wird bei dieser Abschätzung größtenteils der gegenwärtige Gesundheitszustand erfasst. Dieser Effekt wird umso größer, je länger das Erinnerungsintervall ist. Auch sind Erkrankungen mit hoher Symptom-Variabilität besonders anfällig für diese Form der Verzerrung (Recall Bias) [8,9,10]. Im Hinblick auf die Variabilität und Heterogenität von Desmoidtumoren sowie die mediane Beobachtungsdauer im PFS von 19,2 bzw. 10,9 Monaten im Nirogacestat- bzw. Placebo-Arm, scheinen diese Faktoren relevant für die vorliegende Erhebung zu sein. Da für den PGIS keine verwertbaren Auswertungen vorliegen, wird der PGIC, als Zeit bis zu jeglicher Verschlechterung, trotz seiner Limitationen für die Nutzenbewertung herangezogen.

### **Symptomatik mittels EORTC QLQ-C30: Symptomskalen**

Der Endpunkt „EORTC QLQ-C30: Symptomskalen“ wird in der Nutzenbewertung berücksichtigt.

### Operationalisierung

#### *Beschreibung*

Der „Quality of Life Questionnaire – Core Questionnaire“ der „European Organisation for Research and Treatment of Cancer“ (EORTC QLQ-C30) ist ein generisches Instrument zur Erfassung der gesundheitsbezogenen Lebensqualität und von Symptomen bei an Krebs erkrankten Personen. Im Rahmen der Studie DeFi wurde Version 3.0 des EORTC QLQ-C30 verwendet.

Der Fragebogen gliedert sich in Symptomskalen, symptomatische Einzelitems, Funktionsskalen und eine globale Skala zum allgemeinen Gesundheitszustand / zur Lebensqualität.

Im Rahmen der Morbidität werden die folgenden Symptomskalen und Einzelitems berücksichtigt:

- 3 Symptomskalen
  - Fatigue (3 Items)
  - Übelkeit und Erbrechen (2 Items)
  - Schmerzen (2 Items)
- Zusätzliche Einzelitems
  - Dyspnoe
  - Appetitverlust
  - Schlaflosigkeit
  - Obstipation
  - Diarrhö

Nach Angaben in Modul 4 erfolgte die Berechnung der Skalenwerte gemäß dem Scoring Manual des EORTC QLQ-C30, wobei jede Skala in einen Bereich von 0 bis 100 transformiert wurde. Höhere Werte auf Symptomskalen und Einzelitems bedeuten eine stärkere Symptomschwere. In der Endpunktkategorie „Morbidität“ sind die 3 Symptomskalen sowie die Einzelitems zu Symptomen relevant (12 Items).

#### *Bewertung*

Die Operationalisierung ist nachvollziehbar.

#### Patientenrelevanz

Der Endpunkt „EORTC QLQ-C30: Symptomskalen“ wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

#### Validität

Der EORTC QLQ-C30 wurde spezifisch für onkologische Patientengruppen (maligne Erkrankungen) entwickelt und primär in dieser Population validiert. Dies umfasst ein breites Spektrum an Krebserkrankungen in verschiedenen Krankheitsstadien und Behandlungskontexten [3,7]. Jedoch liegen keine Validierungsstudien für den EORTC QLQ-C30 bei benignen Tumorerkrankungen wie Desmoidtumoren vor, was seine Anwendung in diesem Kontext methodisch infrage stellt. Im Gegensatz zu malignen Erkrankungen ist die Symptomatik von Desmoidtumoren primär durch lokale Beschwerden wie Schmerzen und funktionelle Einschränkungen bestimmt, während systemische Krebsfolgen typischerweise nicht vorliegen.

Letztendlich ist aufgrund der fehlenden Untersuchungen im Anwendungsgebiet unklar, inwieweit der EORTC QLQ-C30 die Symptomatik und Lebensqualität umfassend abbildet. Trotz der genannten Limitationen wird der EORTC QLQ C30 zur Nutzenbewertung herangezogen. Die angewendete Responderschwelle von  $\geq 10$  Punkten ist adäquat.

### **2.3.3 Lebensqualität**

#### **EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus**

Der Endpunkt „EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus“ wird in der Nutzenbewertung berücksichtigt.

#### Operationalisierung

##### *Beschreibung*

Für die allgemeinen Ausführungen zum EORTC QLQ-C30 wird auf Kapitel 2.3.2 der vorliegenden Nutzenbewertung verwiesen. Neben den Symptomskalen wurden die 5 Funktionsskalen des EORTC QLQ-C30 sowie die globale Skala „Allgemeiner Gesundheitsstatus / Lebensqualität“ im

Rahmen der Studie DeFi erhoben. Diese wurden als sekundäre Endpunkte operationalisiert. Hierzu zählen die folgenden Skalen:

- 5 Funktionsskalen
  - Physische Funktion (5 Items)
  - Rollenfunktion (2 Items)
  - Emotionale Funktion (4 Items)
  - Kognitive Funktion (2 Items)
  - Soziale Funktion (2 Items)
- Globale Skala „Allgemeiner Gesundheitsstatus / Lebensqualität“ (2 Items)

Die Fragen beziehen sich auf die vergangenen 7 Tage, wobei die Items auf einer Skala von 1 („überhaupt nicht“) bis 4 („sehr“) bewertet wurden. Die Items der globalen Skala werden von 1 („sehr schlecht“) bis 7 („ausgezeichnet“) bewertet. Nach Angaben des pU in Modul 4 erfolgte die Berechnung der Skalenwerte gemäß dem Scoring Manual des EORTC QLQ-C30. Höhere Werte auf den Funktionsskalen sowie der globalen Skala implizieren eine bessere Funktion.

#### *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 Bewertung der Validität des EORTC QLQ-C30 erfolgte im Abschnitt 2.3.2.

### **2.3.4 Sicherheit**

#### **Unerwünschte Ereignisse**

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

#### Operationalisierung

##### *Beschreibung*

Als unerwünschtes Ereignis (UE) werden alle unerwünschten medizinischen Ereignisse definiert, die zeitlich mit der Anwendung der Studienbehandlung in Zusammenhang stehen, unabhängig davon, ob sie in einem kausalen Zusammenhang mit dem verabreichten Arzneimittel stehen oder nicht. Dabei kann ein UE jedes ungünstige und unbeabsichtigte Anzeichen, Symptom oder Krankheit sein, welche/s neu auftritt oder sich verschlimmert und in einem zeitlichen Zusammenhang mit der Gabe der Studienmedikation steht. Dies schließt auch abnormale Laborbefunde mit ein.

UE wurden gemäß „Medical Dictionary for Regulatory Activities“ (MedDRA), Version 24.0 oder neuer, nach Systemorganklasse (SOC) und Preferred Term (PT) kodiert. Die Einteilung der Schweregrade der UE erfolgte mittels „Common Terminology Criteria for Adverse Events“ (CTCAE) des National Cancer Institute (NCI), Version 5.0. Für UE, die nicht in den CTCAE enthalten sind, wird ein alternatives Bewertungssystem verwendet: Grad 1 (mild), Grad 2 (moderat), Grad 3 (schwerwiegend), Grad 4 (lebensbedrohlich), Grad 5 (tödlich).

Ein schwerwiegendes unerwünschtes Ereignis (SUE) ist definiert als ein unerwünschtes medizinisches Ereignis, welches eines der folgenden Kriterien erfüllt: Das Ereignis

- war tödlich,
- war lebensbedrohlich,
- machte eine Hospitalisierung oder eine Verlängerung dieser notwendig,
- führte zu einer anhaltenden Einschränkung oder Beeinträchtigung,

- verursachte eine angeborene Anomalie oder einen Geburtsfehler oder
- wurde als ein anderes wichtiges medizinisches Ereignis eingeschätzt.

Folgende UE von besonderem Interesse (UESI) wurden in den Studienunterlagen präspezifiziert:

- Hautausschlag (wenn klinisch signifikant, Grad 2 und Grad  $\geq 3$ , gemäß CTCAE)
  - Makulopapulöser Ausschlag
  - Pruritischer Ausschlag
  - Erythematöser Ausschlag
  - Follikulitis
  - Hidradenitis suppurativa
- Leberenzyme (Grad  $\geq 2$ , gemäß CTCAE)
  - Aspartat-Aminotransferase erhöht
  - Alanin-Aminotransferase erhöht
  - Alkalische Phosphatase erhöht
- Elektrolytinsuffizienz (Grad  $\geq 3$ , gemäß CTCAE)
  - Hypophosphatämie
  - Hypokaliämie
  - Hypomagnesiämie
- Arzneimittelreaktionen (alle Grade)
  - Allergische Reaktion
  - Anaphylaxie
- Störung des Fortpflanzungssystems (Grad  $\geq 2$ , gemäß CTCAE)
  - Amenorrhö
  - Vorzeitige Menopause / Primäre Ovarialinsuffizienz

Die Erhebung von UE erfolgt vom Zeitpunkt der Unterzeichnung der Einwilligungserklärung bis 30 Tage nach Gabe der letzten Dosis der Studienmedikation. Bei weiblichen Teilnehmenden mit UE in Bezug auf primäre Ovarialinsuffizienz und/oder Amenorrhö wurde der Hormonspiegel alle 3 Monate bis zum Ende des Ereignisses gemessen (oder für mindestens 90 Tage nach Abbruch der Studienbehandlung).

### *Bewertung*

Die Operationalisierung der UE, SUE und UESI ist weitestgehend nachvollziehbar. UE, die mit einer Progression der Erkrankung in Verbindung standen (Zunahme des Schweregrads der untersuchten Krankheit und/oder Zunahme der Krankheitssymptome), sollten nicht als UE gewertet werden. Da keine Auflistung typischer Progressionssymptome vorliegt, ist unklar, ob die Bewertung in den verschiedenen Zentren ähnlich erfolgte. Außerdem ist unklar, wie mit anderen erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung umgegangen wurde. Es ist nicht auszuschließen, dass diese mit in die Erfassung der UE eingingen.

### Patientenrelevanz

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

### Validität

Die Erhebung des Endpunkts „Unerwünschte Ereignisse“ wird als valide eingeschätzt.

### 2.3.5 Erhebungszeitpunkte

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

Studienvisite Endpunkt	Screening <sup>1)</sup>	Baseline <sup>2)</sup>	Zyklus 1	Zyklus 2	Zyklus 3	Jeder weitere Zyklus <sup>3)</sup>	EOT <sup>4)</sup>	Follow-up <sup>5)</sup>
Todesfälle <sup>6)</sup>	x (fortlaufende Erhebung)							
PFS <sup>7)</sup>	-	x (fortlaufende Erhebung) <sup>8)</sup>					x <sup>9)</sup>	-
GODDESS-DTSS <sup>10)</sup> /-DTIS	x	x	x	x	x	x	-	x
BPI-SF <sup>10)</sup>	x	x	x	x	x	x		x
PGIS	x	x	x	x	x	x		x
PGIC	-	-	x	x	x	x		x
EORTC QLQ-C30: Symptomskalen	x	x	x	x	x	x		x
EORTC QLQ-C30: Funktionsskalen u. globaler Gesundheitsstatus	x	x	x	x	x	x		x
Unerwünschte Ereignisse	x (fortlaufende Erhebung)							

<sup>1)</sup> Zeitraum: Bis zu 28 Tage vor Tag 1 (Baseline). Für PRO startet die Erhebung an Tag 1 der Screening-Phase. Für diesen Zeitpunkt erfolgt keine Erhebung des PGIC.

<sup>2)</sup> Zeitraum: Vor Verabreichung der ersten Studienmedikation (bis zu 48 Stunden). Die Erhebung der PRO beginnt 7 Tage vor Zyklus 1 Tag 1.

<sup>3)</sup> Für PRO erfolgt eine fortlaufende monatliche Erhebung vor Beginn des nächsten Zyklus.

<sup>4)</sup> Alle EOT-Visiten sollten vor der Entblindung erfolgen.

<sup>5)</sup> Erfolgt 30 Tage nach der letzten Gabe der Studienbehandlung.

<sup>6)</sup> Todesfälle wurden unter „Sicherheit“ erfasst.

<sup>7)</sup> Primärer Endpunkt (Jegliche qualifizierte klinische Progression oder Tod).

<sup>8)</sup> Im Rahmen der vorliegenden Nutzenbewertung ist für die Operationalisierung des PFS die „klinische Progression“ (jegliche qualifizierte klinische Progression oder Tod) maßgeblich. Hierbei wird entsprechend der Definition der klinischen Progression angenommen, dass die Erhebung fortlaufend erfolgte. Für die radiographische Progression werden ab Zyklus 4 alle 3 Zyklen MRT- oder CT-Aufnahmen zur Tumorbewertung (RECIST v1.1) durchgeführt. Ab Zyklus 7 wird alle 6 Zyklen ein MRT zur Tumorzuberechnung durchgeführt.

<sup>9)</sup> Nur erforderlich, wenn diese nicht innerhalb der letzten 3 Monate durchgeführt wurden.

<sup>10)</sup> Hierbei handelt es sich um einen wöchentlichen Durchschnittsscore. Prinzipiell erfolgt die Erhebung immer 7 Tage vor dem nächsten Zyklus bzw. Follow-up. Zu Baseline und zum Screening erfolgte die Erhebung 7 Tage vor der Visite.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; DTIS: Desmoid Tumor Impact Scale; DTSS: Desmoid Tumor Symptom Scale; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; EOT: End of Treatment; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; MRT: Magnetresonanztomographie; PFS: Progressionsfreies Überleben; PGIC: Patient Global Impression of Change; PGIS: Patient Global Impression of Severity; PRO: Patientenberichteter Endpunkt; RECIST: Response Evaluation Criteria In Solid Tumors.

## 2.4 Statistische Methoden

Grundlage der Auswertung ist der SAP in der Version 2.0 vom 12.05.2022, der auf dem Studienprotokoll (Amendment 5) basiert. Auswertungen zu den PRO basieren auf dem separaten SAP (PRO-Addendum) in der Version 2.0 vom 05.05.2022.

### Datenschnitte

Die Studie ist abgeschlossen (finale Datenbanksperre der OLE-Phase: 19.12.2024). Für die Nutzenbewertung von Nirogacestat liegt der finale Datenschnitt der DB-Phase vom 07.04.2022 vor. Der Datenschnitt war ereignisgesteuert und präspezifiziert für das Erreichen von 51 PFS-Ereignissen; der finale Datenschnitt erfolgte nach bereits 49 von 51 geplanten PFS-Ereignissen. Hierzu wurde in Modul 4 ausgeführt, dass zwischen dem 20.09.2021 und dem 07.04.2022 keine neuen Ereignisse in der Studie DeFi gemeldet wurden. Eine präspezifizierte Interimsanalyse wurde nicht durchgeführt. Für die optionale OLE-Phase wurden keine Daten eingereicht, diese sollen gemäß Studienbericht in einem separaten Bericht dargestellt werden. Der finale Datenschnitt vom 07.04.2022 wird als geeignet für die Nutzenbewertung angesehen.

Für die vorliegende Nutzenbewertung werden die Ergebnisse der DB-Phase herangezogen (Datenschnitt: 07.04.2022).

### Analysepopulationen

- Intention-to-Treat (ITT)-Population: Alle Personen, die in die Studie eingeschlossen und zu einem der beiden Behandlungsarme (Nirogacestat oder Placebo) randomisiert wurden. Personen werden entsprechend der Randomisierung analysiert, unabhängig davon, welche Studienmedikation diese tatsächlich erhalten haben. Die ITT-Population dient als primäre Analysepopulation für Wirksamkeits- und PRO-Analysen.
- Sicherheitspopulation: Alle Personen, die zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben. Die Analyse erfolgt basierend auf der Behandlung, welche die Personen tatsächlich erhalten haben. Die Sicherheitspopulation dient zur primären Analyse der Sicherheitsendpunkte.

### Subgruppenanalysen

Für die relevanten Wirksamkeits- und Sicherheitsendpunkte der Nutzenbewertung (siehe Tabelle 5), ausgenommen „Todesfälle“, wurden verschiedene Subgruppenanalysen präspezifiziert:

- Stratifizierungsfaktor wie bei der Randomisierung angegeben
- Demographische Merkmale
  - o Geschlecht (männlich vs. weiblich)
  - o Alter (nach Quartil)
  - o Abstammung (weiß vs. nicht-weiß)
  - o Geographische Region (Nordamerika vs. Rest der Welt)
  - o BMI (18,5 kg/m<sup>2</sup> vs. 18,5 bis < 25 kg/m<sup>2</sup> vs. 25 bis < 30 kg/m<sup>2</sup> vs. ≥ 30 kg/m<sup>2</sup>)
- Krankheitsmerkmale
  - o Fokale Kategorie (multifokale Erkrankung vs. einzelner Tumor)
  - o Größe der Baseline-Zielläsion nach Quartilen (< 68,90 mm vs. 68,90 bis < 100,35 mm vs. 100,35 bis < 155,35 mm vs. ≥ 155,35 mm)
  - o Lokalisation der Baseline-Zielläsion (intra-abdominal vs. extra-abdominal)

- Vorherige Behandlung
  - o Jegliche vorherige Therapie (ja vs. nein)
  - o Anzahl vorherige Therapielinien (0 vs. 1–3 vs. 4 und mehr)
  - o Vorherige systemische Therapie (ja vs. nein)
  - o Vorheriger chirurgischer Eingriff (ja vs. nein)
  - o Vorherige Strahlenbehandlung (ja vs. nein)
  - o Vorherige Behandlung mit Sorafenib (ja vs. nein)
  - o Vorherige Chemotherapie (ja vs. nein)
  - o Vorherige Behandlung mit Tyrosinkinase-Inhibitoren (ja vs. nein)
  - o Status der Desmoidtumor-Behandlung (behandlungsnaiv, messbar progredient vs. rezidivierend, messbar progredient nach mindestens einer Therapielinie vs. refraktär, messbar progredient nach mindestens einer Therapielinie)
- Genetische Mutation
  - o Anamnese der familiären adenomatösen Polyposis
  - o Vorhandensein einer CTNNB1-Mutation, somatischen CTNNB1-Mutation oder Keimbahn-CTNNB1-Mutation
  - o Vorhandensein einer APC-Mutation, somatischen APC-Mutation oder Keimbahn-APC-Mutation

Zusätzlich wurden im PRO-SAP weitere Subgruppenanalysen für die PRO präspezifiziert:

- Stratifizierungsfaktor wie bei der Randomisierung angegeben
- Baseline Worst Pain Score (unkontrolliert: API > 4 vs. kontrolliert: API ≤ 4)
- Fokale Kategorie (multifokale Erkrankung vs. einzelner Tumor)
- Status der Desmoidtumor-Behandlung (behandlungsnaiv, messbar progredient vs. rezidivierend, messbar progredient nach mindestens einer Therapielinie vs. refraktär, messbar progredient nach mindestens einer Therapielinie)

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

## **Präspezifizierung geplanter und durchgeführter Analysen**

### PRO

Die Endpunkte „GODDESS-DTSS/-DTIS“, „BPI-SF“ sowie „EORTC QLQ-C30“ wurden als sekundäre Endpunkte der Studie DeFi ausgewertet. Der Endpunkt „PGIS“ wurde hingegen explorativ ausgewertet. Die Analysen basieren auf der ITT-Population und Rücklaufquoten (adjustiert und nicht adjustiert) werden für jeden Zeitpunkt der DB-Phase ausgewertet.

### *Kontinuierliche Auswertung*

Für die PRO-Auswertungen (GODDESS-DTSS und -DTIS, BPI-SF, PGIS, EORTC QLQ-C30, PGIC) wurde die Veränderung der PRO-Scores gegenüber Baseline mittels „Mixed Model for Repeated Measures“ auf Basis der „Restricted Maximum Likelihood“-Methode präspezifiziert. Die kontinuierliche Auswertung der Veränderung zu Baseline beschränkt sich auf die On-Treatment-Messzeitpunkte; nicht geplante Visiten sowie End-of-Treatment- und Follow-up-Untersuchungen werden ausgeschlossen. Primäres Ziel der Analyse war die Untersuchung des Behandlungsunterschieds am Ende von Zyklus 10. Diese Auswertungen sind für die vorliegende Nutzenbewertung nicht maßgeblich und sind nicht in Modul 4 berichtet.

### *Responderanalysen*

Es wurden verschiedene Responderanalysen zu den PRO „GODDESS-DTSS“, „GODDESS-DTIS“ sowie „BPI-SF“ (nur API) präspezifiziert. Der Fokus der Analyse bestand darin, Häufigkeiten von Respondern bzw. Non-Respondern zwischen den beiden Behandlungsarmen zum Zeitpunkt Zyklus 10 zu vergleichen. Für den BPI-SF wurde eine literaturbasierte Relevanzschwelle von 2 Punkten (NRS) definiert. Im SAP wurden für die genannten Endpunkte, die im Kontext des GODDESS-Instruments erhoben wurden, verschiedene primäre Relevanzschwellen mit zugehörigen Sensitivitätsschwellen definiert; je nach Score schwanken die Relevanzschwellen zwischen -0,8 und -1,9. Der pU reicht mit dem Dossier für GODDESS-DTSS und -DTIS sowie BPI-SF Auswertungen für eine Relevanzschwelle von 15 % der Skalenspannweite und für den EORTC QLQ-C30 Auswertungen mit einer Relevanzschwelle von 10 Punkten ein. Diese werden für die vorliegende Nutzenbewertung herangezogen.

### Post-hoc-Auswertungen in Modul 4

In Modul 4 wurden post hoc die Ereigniszeitanalysen als „Zeit bis zur Verbesserung bzw. Verschlechterung um  $\geq 15$  % der Skalenspannweite“ (GODDESS-DTSS und -DTIS, BPI-SF) und „Zeit bis zum ersten Auftreten“ (Sicherheit) ausgewertet. Beim PGIC erfolgte die Auswertung der Ereigniszeitanalysen als Zeit bis zu jeglicher erstmaligen Verbesserung bzw. Verschlechterung. Die Analysen erfolgten mittels stratifiziertem Cox-Proportional-Hazards-Modell unter Verwendung der „Exact Method for Ties“, stratifiziert nach primärer Tumorlokalisierung (intra-abdominal vs. extra-abdominal) bei Randomisierung. Für die Überprüfung der Annahme proportionaler Hazard-Raten ( $\alpha$ -Niveau = 0,05) im Cox-Modell liegen keine Angaben in Modul 4 vor, welche spezifischen Testverfahren angewendet wurden. Im Falle einer Verletzung der Proportional-Hazard-Annahme wurde das relative Risiko (RR) berechnet.

Personen, bei denen zum Zeitpunkt der Analyse keine Symptomverschlechterung beobachtet wurde, werden zum Zeitpunkt der letzten verfügbaren PRO-Bewertung (d. h. zum Datum des letzten nicht fehlenden Werts) oder vor dem Datenschnitt der Analysen zensiert. Basierend hierauf ist anzunehmen, dass Daten aus der End-of-Treatment- sowie Follow-up-Untersuchung in die Ereigniszeitanalysen eingingen. Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha$ -Niveau = 0,025). Für den PGIS wurden keine Ereigniszeitanalysen durchgeführt. Für die Sicherheitsendpunkte und PRO wurden Daten zensiert, die nach Wechsel des Behandlungsarms zu Nirogacestat in der OLE-Phase erfasst wurden.

### **Fehlende Werte und Daten-Imputation (nur für PRO geplant)**

Für die PRO „GODDESS-DTSS“ und „GODDESS-DTIS“, „BPI-SF“ sowie „EORTC QLQ-C30“ wurden Sensitivitätsanalysen in der ITT-Population präspezifiziert, in denen fehlende Daten mittels multipler Imputation imputiert wurden. Hierbei wurden „Pattern-mixture Models“ verwendet, um die Möglichkeit der Annahme „Missing Not at Random“ zu prüfen. Außerhalb von Sensitivitätsanalysen und PRO wurden keine Imputationen durchgeführt. Sensitivitätsanalysen für die Ereigniszeitanalysen liegen nicht vor.

### Einschätzung der statistischen Auswertungen

Aufgrund unterschiedlicher Behandlungszeiten in den Studienarmen (im Median 20,6 vs. 11,4 Monate; siehe Tabelle 9) werden für alle PRO und Sicherheitsendpunkte die Ereigniszeitanalysen als die relevante Auswertung eingeschätzt. Desmoidtumore sind lokal aggressiv wachsende, nicht metastasierende Tumore der Weichteile, deren klinische Morbidität primär auf ihr infiltratives und potentiell progredientes Wachstum zurückzuführen ist. Dies kann zu irreversiblen Funktionsdefiziten, neurovaskulären Kompressionen, starken Schmerzen und erheblichen Einschränkungen der Lebensqualität führen. Gemäß Anwendungsgebiet ist die Zielpopulation von Nirogacestat von fortschreitenden Desmoidtumoren betroffen. Aufgrund des im vorliegenden Anwendungsgebiet weiterhin zu erwartenden progredienten Krankheitsverlaufs, wird die Zeit bis zur Verschlechterung als relevante Auswertung für die Nutzenbewertung angesehen. Die p-Werte stammen aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ). Bei nicht berechenbaren Hazard Ratios (keine Ereignisse in einem Arm) oder Verletzung der Proportional-Hazard-Annahme wurden RR mit einem Cochran-Mantel-Haenszel-Modell berechnet. Aufgrund unterschiedlicher Beobachtungszeiten ist eine Effektschätzung anhand des RR jedoch nicht sachgerecht, sodass auf die Darstellung verzichtet wird.

Für den PGIS liegen anders als bei allen Endpunkten keine Ereigniszeitanalysen vor. Die Post-hoc-Analyse zur Veränderung des PGIS zu Zyklus 24 gegenüber Baseline wird aufgrund der geringen Rücklaufquote zu Zyklus 24 als nicht adäquat bewertet und daher nicht in der vorliegenden Nutzenbewertung dargestellt.

## 2.5 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 7: Verzerrungspotential der Studie DeFi

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppeneinteilung	Verblindung Patientin/Patient	Verblindung Behandlungsperson	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
DeFi	Ja <sup>1)</sup>	Ja	Unklar <sup>2)</sup>	Unklar <sup>2)</sup>	Nein	Nein	Niedrig

<sup>1)</sup> Die Randomisierung erfolgte mittels IRT stratifiziert nach der Tumorlokalisierung (intra-abdominal vs. extra-abdominal).

<sup>2)</sup> Aufgrund spezifischer charakteristischer UE wie bspw. Diarrhö, Alopezie, vorzeitige Menopause und verschiedene Formen des Hautausschlags kann eine potentielle Entblindung nicht ausgeschlossen werden.

Abkürzungen: IRT: Interactive Response Technology; UE: Unerwünschtes Ereignis.

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 DeFi**

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
<b>Morbidität</b>					
PFS <sup>1)</sup>	Ja	Ja	Nein	Nein	Niedrig
GODDESS-DTSS/-DTIS	Unklar <sup>2)</sup>	Nein <sup>3)</sup>	Nein	Ja <sup>4)</sup>	Hoch
BPI-SF	Unklar <sup>2)</sup>	Nein <sup>3)</sup>	Nein	Ja <sup>4)</sup>	Hoch
PGIC	Unklar <sup>2)</sup>	Nein <sup>3)</sup>	Nein	Ja <sup>4)</sup>	Hoch
EORTC QLQ-C30: Symptomskalen	Unklar <sup>2)</sup>	Nein <sup>3)</sup>	Nein	Ja <sup>4)</sup>	Hoch
<b>Lebensqualität</b>					
EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus	Unklar <sup>2)</sup>	Nein <sup>3)</sup>	Nein	Ja <sup>4)</sup>	Hoch
<b>Sicherheit</b>					
Schwere UE (Gesamtrate, SOC/PT)	Ja	Ja	Nein	Ja <sup>4)</sup>	Hoch
SUE (Gesamtrate, SOC/PT)	Ja	Ja	Nein	Ja <sup>4)</sup>	Hoch
Abbruch wegen UE (Gesamtrate)	Unklar <sup>2)</sup>	Ja	Nein	Ja <sup>4)</sup>	Hoch
UE, jeglichen Schweregrads (SOC/PT)	Unklar <sup>2)</sup>	Ja	Nein	Ja <sup>4)</sup>	Hoch

<sup>1)</sup> Der Endpunkt wird als primärer Endpunkt der Studie ergänzend in der Nutzenbewertung dargestellt.

<sup>2)</sup> Aufgrund spezifischer charakteristischer UE wie bspw. Diarrhö, Alopezie und verschiedene Formen des Hautausschlags kann eine potentielle Entblindung nicht ausgeschlossen werden.

<sup>3)</sup> Hoher Anteil fehlender Werte. Die Rücklaufquoten sind bereits zu bzw. ab Zyklus 4 < 70 %. Es fehlen Angaben zu Zensierungsgründen.

<sup>4)</sup> Bei frühzeitigem Behandlungs-/Studienabbruch wurden die Personen lediglich 30 Tage nachbeobachtet. Dies führt zu einer verkürzten Erhebung und einer potentiellen Verzerrung durch informative Zensierung.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; DTIS: Desmoid Tumor Impact Scale; DTSS: Desmoid Tumor Symptom Scale; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; ITT: Intention to Treat; PFS: Progressionsfreies Überleben; PGIC: Patient Global Impression of Change; PT: Preferred Term; SOC: Systemorganklasse; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

### 3 Ergebnisse der eingeschlossenen Studie

Es werden die Ergebnisse der DB-Phase der Studie DeFi zum Datenschnitt (Ende der DB-Phase) vom 07.04.2022 dargestellt.

#### 3.1 Studiencharakteristika und Studienmedikation

Es wurden insgesamt 201 Personen gescreent, von denen 142 in die Studie eingeschlossen und randomisiert wurden: 70 Personen in den Nirogacestat-Arm, 72 Personen in den Placebo-Arm. Angaben zur Beobachtungsdauer liegen nicht vor. Es wird aber davon ausgegangen, dass sich die Unterschiede zwischen den Behandlungsarmen bezogen auf die mediane Behandlungsdauer (20,6 vs. 11,4 Monate) auch in der Beobachtungsdauer widerspiegelt.

Tabelle 9: Allgemeine Angaben zur DB-Phase; Studie DeFi (Datenschnitt: 07.04.2022)

Studie DeFi Allgemeine Angaben	Nirogacestat	Placebo
ITT-Population <sup>1)</sup> , n (%)	70 (100)	72 (100)
Sicherheitspopulation <sup>1)</sup> , n (%)	69 (99)	72 (100)
<i>Behandlungsstatus während der DB-Phase, n (%)</i>		
Primärer Endpunkt erreicht <sup>2)</sup>	12 (17)	35 (49)
Vorzeitiger Behandlungsabbruch	21 (30)	14 (19)
Aufgrund von UE	14 (20)	1 (1)
Klinische Progression (unqualifiziert)	1 (1)	1 (1)
Fehlende Compliance	1 (1)	1 (1)
Andere Gründe	5 (7)	11 (15)
Weiterhin unter Behandlung	36 (51)	23 (32)
Personen, die die DB-Phase vorzeitig abgebrochen haben, n (%)		
Aufgrund von UE	11 (16)	1 (1)
Entscheidung Patientin/Patient	4 (6)	7 (10)
Radiographische Krankheitsprogression	7 (10)	1 (1)
Klinische Progression (qualifiziert) <sup>3)</sup>	1 (1)	6 (8)
Klinische Progression (unqualifiziert)	1 (1)	1 (1)
Entscheidung Prüfpersonal	2 (3)	0 (0)
Tod	0 (0)	1 (1)
Lost to Follow-up	0 (0)	1 (1)
Andere Gründe	4 (6)	3 (4)
Beobachtungsdauer in Monaten, Median (min; max)	k. A.	k. A.
Behandlungsdauer <sup>4)</sup> in Monaten, Median (min; max)	20,6 (0,3; 33,6)	11,4 (0,2; 32,5)
Personen, die in die offene OLE-Phase übergehen konnten <sup>5)</sup> , n (%)	11 (16)	29 (40)
Personen, die in die OLE-Phase eingeschlossen wurden	4 (36)	28 (97)
Personen, die in die OLE-Phase nicht eingeschlossen wurden <sup>6)</sup>	7 (64)	1 (3)

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

<sup>2)</sup> Anzahl der Progressionsereignisse bezog sich auf die Gesamtereignisse beim Abbruch der Studienbehandlung, nicht auf die Analyse des PFS.

<sup>3)</sup> Bestätigung der qualifizierten klinischen Progression durch ein unabhängiges, verblindetes, zentrales EAC

<sup>4)</sup> Behandlungsdauer = (Datum der letzten Dosis – Datum der ersten Dosis + 1) / 30,4375.

<sup>5)</sup> Person war in die DB-Phase eingeschlossen und die primäre PFS-Analyse ist abgeschlossen oder Person im Placebo-Arm hat gemäß zentraler Bildauswertung eine radiographisch bestätigte fortschreitende Erkrankung (gemäß RECIST v1.1) oder Person im Nirogacestat-Arm hat gemäß zentraler Bildauswertung eine radiographisch bestätigte fortschreitende Erkrankung (gemäß RECIST v1.1), aber Person hat weiterhin klinischen Nutzen ohne signifikante Toxizität (Einschätzung

Prüfpersonal) oder adäquate Organ- und Knochenmarkfunktion, wie in Einschlusskriterien der RCT-Phase beschrieben (basierend auf Hämatologie- u. Serumchemie-Ergebnissen innerhalb von 14 Tagen vor Einschreibung in OLE-Phase).  
<sup>6</sup> Gründe liegen nicht vor.

Abkürzungen: DB: doppelblind; EAC: Endpoint Adjudication Committee; ITT: Intention to Treat; k. A.: keine Angabe; OLE: Open-Label-Extension; PFS: Progressionsfreies Überleben; RCT: Randomized Controlled Trial; RECIST: Response Evaluation Criteria In Solid Tumors; UE: Unerwünschtes Ereignis.

## Demographische und krankheitsspezifische Charakteristika

Tabelle 10: Charakterisierung der Studienpopulation; Studie DeFi (DB-Phase), ITT-Population  
(Datenschnitt: 07.04.2022)

Studie DeFi Charakterisierung der Studienpopulation	Nirogacestat N = 70	Placebo N = 72
<i>Alter (Jahre) bei Einwilligung</i> MW (SD) Median (min; max)	37,5 (14,4) 33,5 (18; 73)	37,0 (34,5) 34,5 (18; 76)
<i>Altersgruppe (Jahre), n (%)</i> < 27 27 bis < 34 34 bis < 46 ≥ 46	20 (29) 15 (21) 13 (19) 22 (31)	14 (19) 18 (25) 25 (35) 15 (21)
<i>Geschlecht, n (%)</i> männlich weiblich	25 (36) 45 (64)	25 (35) 47 (65)
<i>Abstammung, n (%)</i> weiß schwarz oder afroamerikanisch asiatisch andere	64 (91) 4 (6) 1 (1) 1 (1)	54 (75) 5 (7) 3 (4) 10 (14)
<i>Region, n (%)</i> Nordamerika Europa <sup>1)</sup>	44 (63) 26 (37)	53 (74) 19 (26)
<i>BMI (kg/m<sup>2</sup>)<sup>2)</sup></i> n (%) MW (SD) Median (min; max)	69 (99) 26,3 (6,94) 23,8 (17,7; 52,3)	71 (99) 27,1 (6,12) 25,7 (17,5; 44,8)
<i>BMI-Gruppe (kg/m<sup>2</sup>), n (%)</i> < 18,5 18,5 bis < 25 25 bis < 30 ≥ 30	3 (4) 36 (51) 15 (21) 15 (21)	2 (3) 28 (39) 23 (32) 18 (25)
<i>ECOG-Status zu Baseline<sup>3)</sup>, n (%)</i> 0 1 2	51 (73) 18 (26) 1 (1)	52 (72) 20 (28) 0 (0)

<b>Studie DeFi Charakterisierung der Studienpopulation</b>	<b>Nirogacestat N = 70</b>	<b>Placebo N = 72</b>
<b>Krankheitsmerkmale</b>		
<i>Symptomatische Belastung zu Baseline mittels gewichtetem DTSS-Gesamtscore<sup>4)</sup></i>		
n (%)	70 (100)	71 (99)
MW (SD)	3,3 (2,33)	3,5 (2,57)
Median (min; max)	3,2 (0; 8)	3,2 (0; 9)
<i>Zeit seit Diagnose bis Randomisierung<sup>5)</sup> (Monate)</i>		
MW (SD)	59,3 (72,9)	61,7 (77,0)
Median (min; max)	30,2 (0,7; 307,2)	31,2 (3,4; 343,1)
<i>Randomisierungsstrata, n (%)</i>		
intra-abdominal	17 (24)	18 (25)
extra-abdominal	53 (76)	54 (75)
<i>DT-Behandlungsstatus, n (%)</i>		
behandlungsnaiv	18 (26)	14 (19)
refraktär	43 (61)	55 (76)
rezidivierend	9 (13)	3 (4)
<i>Fokale Kategorie, n (%)</i>		
einzel	43 (61)	41 (57)
multifokal	27 (39)	31 (43)
<i>Anzahl der Zieltumore<sup>6)</sup></i>		
n (%)	69 (99)	72 (99)
MW (SD)	1,3 (0,64)	1,6 (0,93)
Median (min; max)	1,0 (1; 4)	1,0 (1; 5)
1	50 (71)	45 (63)
2	15 (21)	18 (25)
3	3 (4)	4 (6)
4	1 (1)	4 (6)
5	0 (0)	1 (1)
<i>Lokalisierung der Zieltumore, n (%)</i>		
Bauchwand	12 (17)	19 (26)
Brustwand	9 (13)	9 (13)
Nacken und Kopf	8 (11)	6 (8)
Untere Extremitäten	14 (20)	11 (15)
Mesenterium und Becken	16 (23)	16 (22)
Paraspinal	6 (9)	8 (11)
Obere Extremitäten	10 (14)	13 (18)
Andere	4 (6)	8 (11)
<i>Größe der Zieltumore zu Baseline gemäß RECIST (mm)<sup>7)</sup></i>		
n (%)	69 (99)	72 (100)
MW (SD)	111,2 (70,6)	121,2 (66,3)
Median (min; max)	91,6 (22,3; 356,2)	115,7 (19,9; 400,9)

<b>Studie DeFi Charakterisierung der Studienpopulation</b>	<b>Nirogacestat N = 70</b>	<b>Placebo N = 72</b>
<i>Anzahl der Nicht-Zieltumore<sup>6)</sup></i>		
MW (SD)	0,6 (0,96)	1,0 (1,57)
Median (min; max)	0 (0; 4)	0 (0; 7)
0	47 (67)	37 (51)
1	11 (16)	19 (26)
2	8 (11)	4 (6)
3	3 (4)	7 (10)
4	1 (1)	2 (3)
5	0 (0)	1 (1)
6	0 (0)	0 (0)
7	0 (0)	2 (3)
<i>Lokalisierung der Nicht-Zieltumore, n (%)</i>		
Bauchwand	0 (0)	4 (6)
Brustwand	1 (1)	4 (6)
Untere Extremitäten	8 (11)	6 (8)
Mesenterium und Becken	2 (3)	3 (4)
Obere Extremitäten	3 (4)	3 (4)
Andere	5 (7)	5 (7)
<i>Anamnese von FAP, n (%)</i>		
ja	11 (16)	13 (18)
nein	59 (84)	59 (82)
<i>Jegliche Mutation, n (%)</i>		
ja	54 (77)	58 (81)
nein	15 (21)	13 (18)
unbekannt	1 (1)	1 (1)
<i>Jegliche somatische Mutation, n (%)</i>		
ja	52 (74)	53 (74)
nein	0 (0)	1 (1)
unbekannt	18 (26)	18 (25)
<i>Jegliche Keimbahnmutation, n (%)</i>		
ja	9 (13)	12 (17)
nein	59 (84)	59 (82)
unbekannt	2 (3)	1 (1)
<i>Jegliche APC-Mutation, n (%)</i>		
ja	13 (19)	16 (22)
nein	56 (80)	55 (76)
unbekannt	1 (1)	1 (1)
<i>Jegliche somatische APC-Mutation, n (%)</i>		
ja	11 (16)	11 (15)
nein	41 (59)	43 (60)
unbekannt	18 (26)	18 (25)
<i>Jegliche Keimbahn-APC-Mutation, n (%)</i>		
ja	9 (13)	12 (17)
nein	59 (84)	59 (82)
unbekannt	2 (3)	1 (1)

<b>Studie DeFi Charakterisierung der Studienpopulation</b>	<b>Nirogacestat N = 70</b>	<b>Placebo N = 72</b>
<i>Jegliche CTNNB1-Mutation, n (%)</i>		
ja	43 (61)	42 (58)
nein	26 (37)	29 (40)
unbekannt	1 (1)	1 (1)
<i>Jegliche somatische CTNNB1-Mutation, n (%)</i>		
ja	43 (61)	42 (58)
nein	9 (13)	12 (17)
unbekannt	18 (26)	18 (25)
<i>Jegliche Keimbahn-CTNNB1-Mutation, n (%)</i>		
ja	0 (0)	0 (0)
nein	68 (97)	71 (99)
unbekannt	2 (3)	1 (1)
<b>Vorherige DT-Therapien</b>		
<i>Jegliche Vortherapie (systemisch, Bestrahlung, Operation), n (%)</i>		
MW (SD)	3,0 (3,5)	3,5 (3,8)
Median (min; max)	2,0 (0; 14)	2,0 (0; 19)
ja	52 (74)	58 (81)
nein	18 (26)	14 (19)
<i>Bestrahlung Vortherapie, n (%)</i>		
ja	16 (23)	16 (22)
nein	54 (77)	56 (78)
<i>Operation Vortherapie, n (%)</i>		
ja	31 (44)	44 (61)
nein	39 (56)	28 (39)
<i>Systemische Vortherapie, n (%)</i>		
ja	43 (61)	44 (61)
MW (SD)	1,6 (2,0)	1,7 (2,1)
Median (min; max)	1,0 (0; 9)	1,0 (0; 11)
nein	27 (39)	28 (39)
<i>Vorherige Sorafenib-Exposition, n (%)</i>		
ja	17 (24)	18 (25)
nein	53 (76)	54 (75)

<sup>1)</sup> Aus Deutschland nahmen 7 Personen (10 %) im Nirogacestat-Arm und 3 Personen (4 %) im Placebo-Arm teil.

<sup>2)</sup> Baseline war definiert als die letzte Messung vor der ersten Verabreichung der Studienbehandlung. Der Baseline-BMI wurde wie folgt berechnet: Baseline-Gewicht in kg / (Baseline-Größe in m)<sup>2</sup>.

<sup>3)</sup> Baseline war definiert als die letzte Messung vor der ersten Verabreichung der Studienbehandlung.

<sup>4)</sup> Die Items werden auf einer 11-stufigen numerischen Bewertungsskala von 0 („nicht vorhanden“) bis 10 („so schlimm, wie nur vorstellbar“) bewertet. Höhere Werte implizieren eine stärkere Symptomschwere.

<sup>5)</sup> Zeit von der Diagnose bis zur Randomisierung wurde wie folgt berechnet: (Datum der Randomisierung – Diagnosedatum + 1) / 30,4375. Diagnosedaten wurden bei Bedarf imputiert.

<sup>6)</sup> Die Anzahl der Zieltumore und Nicht-Zieltumore sowie die Lokalisation(en) der Tumore wurden von den zentralen Reviewern bewertet. Zieltumore waren dabei die größten messbaren Tumore, während Nicht-Zieltumore alle anderen vorhandenen läSIONalen Strukturen waren.

<sup>7)</sup> Die Baseline-Zieltumor-Größe wurde berechnet, indem die längsten Durchmesser für jeden Tumor und jeden Reviewer summiert und anschließend der Durchschnitt dieser Summen über alle Reviewer gebildet wurde.

Abkürzungen: APC: Adenomatous Polyposis Coli; CTNNB1: Catenin Beta-1; DT: Desmoidtumor; DTSS: Desmoid Tumor Symptom Scale; ECOG: Eastern Cooperative Oncology Group; FAP: Familiäre adenomatöse Polyposis; ITT: Intention to Treat; MW: Mittelwert; RECIST: Response Evaluation Criteria In Solid Tumors; SD: Standardabweichung.

## Exposition mit der Studienmedikation

Tabelle 11: Angaben zur Exposition mit der Studienmedikation; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022)

Studie DeFi Exposition mit der Studienmedikation	Nirogacestat N = 69	Placebo N = 72
Dauer der Exposition (Monate) <sup>1)</sup>		
MW (SD)	17,4 (9,92)	13,3 (8,44)
Median (min; max)	20,6 (0,3; 33,6)	11,4 (0,2; 32,5)
Anzahl an angetretenen Zyklen, n (%)		
Zyklus 1	67 (97)	71 (99)
Zyklus 2	61 (88)	71 (99)
Zyklus 3	59 (86)	67 (93)
Zyklus 4	58 (84)	62 (86)
Zyklus 5	56 (81)	58 (81)
Zyklus 6 bis < Zyklus 13	56 (81)	56 (78)
Zyklus 13 und länger	45 (65)	34 (47)
Dosisanpassungen, n (%)	45 (65)	12 (17)
Unterbrechung	45 (65)	12 (17)
Reduktion	29 (42)	0 (0)

<sup>1)</sup> Die Expositionsdauer wurde wie folgt berechnet: (Datum der letzten Dosis – Datum der ersten Dosis + 1) / 30,4375. Abkürzungen DB: doppelblind; MW: Mittelwert; SD: Standardabweichung.

## Protokollabweichungen

Insgesamt wiesen 38 Personen (54 %) im Nirogacestat-Arm und 47 Personen (65 %) im Placebo-Arm mindestens eine Protokollabweichung auf. Relevante Protokollabweichungen wurden bei jeweils 4 Personen (6 %) im Nirogacestat- und Placebo-Arm gemeldet. Diese Abweichungen umfassten die Fehlstratifizierung aufgrund falscher Tumorlokalisierung sowie die Aufnahme von Personen, die die Ein-/Ausschlusskriterien nicht erfüllten. Anhand der in der Studie DeFi dokumentierten Protokollverletzungen ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

## Begleitmedikation

In der Studie erhielten 100 % der Personen im Nirogacestat- und 97 % der Personen im Placebo-Arm mindestens eine Begleitmedikation (siehe Tabelle 12). Begleitmedikamente wurden gezählt von der ersten Einnahme der Studienbehandlung (oder wenn diese vor Studienbehandlung eingenommen wurden und über den Beginn der Studienbehandlung hinaus) bis zur letzten Studienbehandlung + 30 Tage.

Es zeigt sich, dass die Mehrzahl der in Tabelle 12 genannten Begleitmedikationen häufiger im Nirogacestat- als im Placebo-Arm verabreicht wurden. Angaben zur Beobachtungsdauer für die Begleitmedikation liegen nicht vor. Es wird aber davon ausgegangen, dass sich die Unterschiede zwischen den Behandlungsarmen bezogen auf die Behandlungsdauer (20,6 vs. 11,4 Monate) auch in der Beobachtungsdauer für Begleitmedikation widerspiegelt. Es ist daher unklar, ob die häufigere Verabreichung im Interventionsarm auf die längere Beobachtungsdauer oder Unterschiede in der Notwendigkeit für Begleitmedikation zurückzuführen sind. Aufgrund des Nebenwirkungsprofils erscheint es plausibel, dass bspw. Antiemetika häufiger im Interventionsarm benötigt wurden.

*Tabelle 12: Begleitmedikation bei mindestens 15 % der Studienpopulation; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022)*

<b>Studie DeFi Begleitmedikationen</b> <i>ATC-Klasse Wirkstoffname</i>	<b>Nirogacestat</b> <b>N = 69</b> <i>n (%)</i>	<b>Placebo</b> <b>N = 72</b> <i>n (%)</i>
Jegliche Begleitmedikation <sup>1)</sup>	69 (100)	70 (97)
Antipropulsiva	37 (54)	10 (14)
Loperamid	36 (52)	9 (13)
Andere Virusimpfstoffe	36 (52)	14 (19)
Tozinameran	22 (32)	9 (13)
COVID-19-Impfstoff mRNA (mRNA-1273)	13 (19)	6 (8)
Anilide	32 (46)	35 (49)
Paracetamol	32 (46)	34 (47)
Serotonin-(5HT3)-Antagonisten	27 (39)	21 (29)
Ondansetron	26 (38)	21 (29)
Propionsäure-Derivate	24 (35)	20 (28)
Ibuprofen	21 (30)	17 (24)
Glukokortikoide	23 (33)	12 (17)
Natürliche Opiumalkaloide	23 (33)	24 (33)
Oxycodon	14 (20)	19 (26)
Tetracycline	21 (30)	4 (6)
Doxycyclin	11 (16)	3 (4)
Andere Mineralstoffe	18 (26)	3 (4)
Amide	17 (25)	13 (18)
Lidocain	15 (22)	9 (13)
Andere Mittel zur lokalen oralen Behandlung	16 (23)	1 (1)
Andere Analgetika und Antipyretika	16 (23)	20 (28)
Gabapentin	11 (16)	11 (15)
Andere Antiemetika	15 (22)	6 (8)
Protonenpumpenhemmer	15 (22)	15 (21)
Benzodiazepin-Derivate	14 (20)	17 (24)
Lorazepam	11 (16)	6 (8)
Kombinationen von Penicillinen, einschl. Beta-Lactamase-Hemmer	14 (20)	4 (6)
Amoxicillin-Trihydrat; Kaliumclavulanat	13 (19)	3 (4)
Antiinfektiva zur Behandlung von Akne	13 (19)	3 (4)
Kalium	13 (19)	2 (3)
Vitamin D und Analoga	13 (19)	10 (14)
Opioide in Kombination mit Nicht-Opioid-Analgetika	12 (17)	10 (14)
Kortikosteroide, schwach (Gruppe I)	11 (16)	1 (1)
Hydrocortison	11 (16)	1 (1)

<b>Studie DeFi</b> <b>Begleitmedikationen</b> <i>ATC-Klasse</i> <i>Wirkstoffname</i>	<b>Nirogacestat</b> <b>N = 69</b> <i>n (%)</i>	<b>Placebo</b> <b>N = 72</b> <i>n (%)</i>
Osmotisch wirkende Laxantien	11 (16)	10 (14)
Andere Antidepressiva	11 (16)	6 (8)
Andere zentral wirkende Mittel	11 (16)	5 (7)
Andere Opioide	11 (16)	13 (18)
Tramadol	7 (10)	13 (18)

<sup>1)</sup> Begleitmedikation umfasst alle Medikamente, die zeitgleich mit der Studienbehandlung eingenommen wurden. Dies schloss Arzneimittel ein, die vor der ersten Dosis der Studienbehandlung begonnen und nach der ersten Dosis fortgesetzt wurden bis zur letzten Studienbehandlung + 30 Tage.

Abkürzungen: ATC: Anatomisch-therapeutisch-chemisches Klassifikationssystem; DB: doppelblind.

### 3.2 Mortalität

In der Studie DeFi wurden Todesfälle kontinuierlich im Rahmen der Sicherheitserfassung erhoben. Während der DB-Phase, bzw. bis zum Datenschnitt vom 07.04.2022, wurde ein Todesfall im Placebo-Arm berichtet.

### 3.3 Morbidität

#### PFS (ergänzend dargestellt)

Der Endpunkt „PFS“ wird in der folgenden Operationalisierung als klinische Progression in der Nutzenbewertung ergänzend dargestellt. Die mediane Follow-up-Dauer betrug 19,2 Monate (Spannweite 0–31) im Nirogacestat-Arm und 10,9 Monate (Spannweite 0–31) im Placebo-Arm. Die Auswertungen für die ITT-Population werden in Tabelle 13 und Abbildung 1 ergänzend dargestellt. Zensierungen und Zensierungsgründe wurden nicht in Modul 4 aufgeführt.

Obwohl das Hazard Ratio (HR) von 0,14 mit einem einseitigen p-Wert von 0,014 auf einen starken positiven Behandlungseffekt hindeutet, ist diese Schätzung aufgrund der geringen Anzahl an Ereignissen statistisch noch nicht robust. Das breite Konfidenzintervall (KI) [0,02; 1,10] – welches den Wert 1 (kein Effekt) einschließt – unterstreicht die erhebliche Unsicherheit der Schätzung. Bei niedriger Ereignisanzahl kann bereits ein einzelnes zusätzliches Ereignis das HR und den p-Wert maßgeblich verändern.

Tabelle 13: Ergebnisse zum PFS (ergänzend dargestellt); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

Studie DeFi PFS (ergänzend dargestellt) <sup>1)</sup>	Nirogacestat N = 70	Placebo N = 72
Ereignis, n (%)	1 (1)	7 (10)
Tod	0 (0)	1 (1)
Klinische Progression <sup>2)</sup>	1 (1)	6 (8)
Zensierungen	k. A.	k. A.
Mediane Zeit bis zum Ereignis, Monate [95%-KI]	n. a. [n. a.; n. a.]	n. a. [n. a.; n. a.]
HR [95%-KI] <sup>3)</sup>		0,14 [0,02; 1,10]
p-Wert <sup>4)</sup>		0,014

<sup>1)</sup> Definiert als jegliche qualifizierte klinische Progression oder Tod.

<sup>2)</sup> Die Beurteilung einer klinischen Progression erfolgt durch das Prüfpersonal. Die klinische Progression wurde als das Auftreten oder die Verschlechterung von Symptomen definiert, die zu einer globalen Verschlechterung des Gesundheitszustands führten und das Ende der Studienbehandlung und den Beginn einer anschließenden Therapie (Strahlentherapie, Operation oder systemische Therapie einschl. Chemotherapie oder Tyrosinkinase-Inhibitoren) zur Folge hatten. Die Überprüfung und Bestätigung der Progressionsereignisse erfolgte durch das verblindete, zentrale EAC.

<sup>3)</sup> Das HR wurde aus dem stratifizierten Cox-Proportional-Hazards-Modell geschätzt. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgt nach der Tumorlokalisation.

<sup>4)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen.

Abkürzungen: DB: doppelblind; EAC: Endpoint Adjudication Committee; HR: Hazard Ratio; ITT: Intention to Treat; k. A.: keine Angabe; KI: Konfidenzintervall; n. a.: nicht anwendbar.

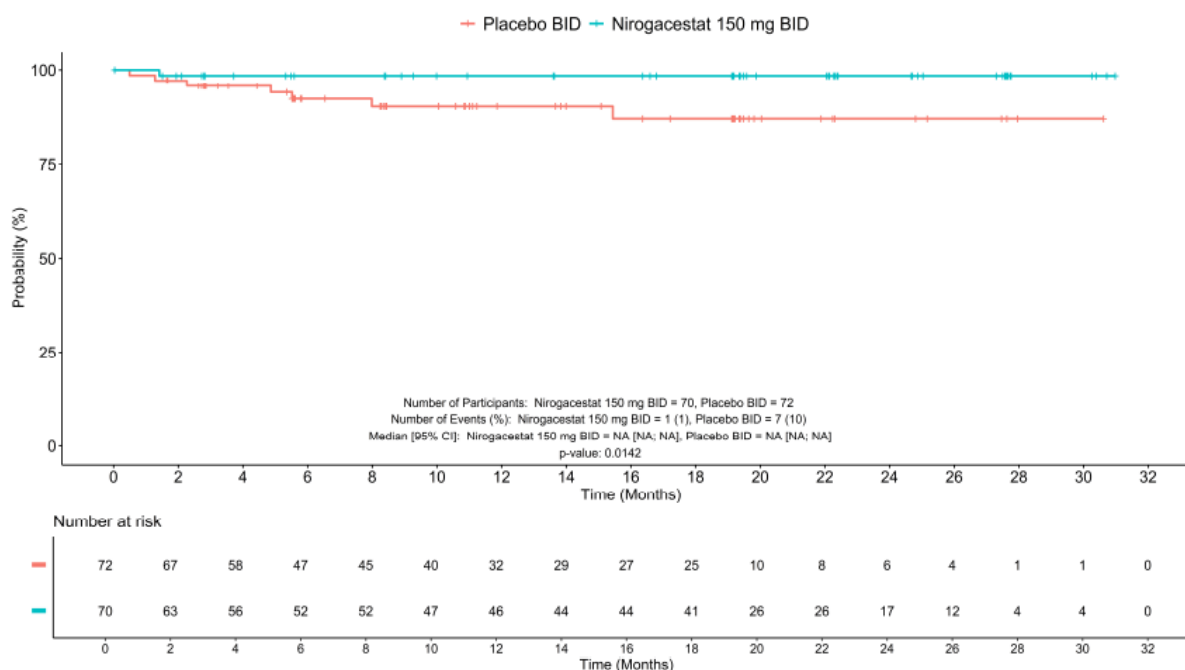


Abbildung 1: Kaplan-Meier-Kurve für PFS (Zeit bis zu jeglicher qualifizierten klinischen Progression oder zum Tod); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

## Symptomatik mittels GODDESS-DTSS/-DTIS

### GODDES-DTSS

#### Responderanalyse

Für die GODDESS-DTSS werden in der vorliegenden Nutzenbewertung der „Gewichtete DTSS-Gesamtscore“ und der Domänenscore „Intra-abdominale Symptome“ berücksichtigt.

Bei der Responderanalyse „Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite“ zeigte sich ein signifikanter Unterschied ( $p = 0,004$ ) zwischen den beiden Behandlungsarmen für den gewichteten DTSS-Gesamtscore zugunsten von Nirogacestat (Tabelle 14, Abbildung 2).

*Tabelle 14: GODDESS-DTSS – Responderanalyse (Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)*

Studie DeFi GODDESS-DTSS <sup>1)</sup>	Nirogacestat N = 70	Placebo N = 72	HR [95%-KI] <sup>2)</sup> ; p-Wert <sup>3)</sup>
<b>Gewichtete DTSS-Gesamtscore</b>			
Ereignisse (%)	26 (37)	40 (56)	
Mediane Zeit (Monate) [95%-KI]	n. a. [10,9; n. a.]	4,40 [2,37; n. a.]	0,50 [0,30; 0,84]; 0,004
<b>Subdomänenscore „Schmerz“</b>			
Ereignisse (%)	21 (30)	37 (51)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	5,32 [2,56; n. a.]	0,43 [0,25; 0,75]; -
<b>Subdomänenscore „Extra-abdominale Symptome“</b>			
Ereignisse (%)	22 (31)	39 (54)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	5,36 [2,66; n. a.]	0,44 [0,26; 0,74]; -
<b>Domänenscore „Intra-abdominale Symptome“</b>			
Ereignisse (%)	12 (17)	15 (21)	
Mediane Zeit (Monate) [95%-KI]	n. a. [1,64; n. a.]	3,52 [0,76; n. a.]	0,95 [0,41; 2,22] 0,45

<sup>1)</sup> Eine Zensierung erfolgte bei Personen, bei denen zum Zeitpunkt der Analyse keine Symptomverschlechterung beobachtet wurde. Diese wurden zum Zeitpunkt der letzten verfügbaren PRO-Bewertung zum oder vor dem Datenschnitt der Analysen zensiert.

<sup>2)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>3)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen.

Abkürzungen: DB: doppelblind; DTSS: Desmoid Tumor Symptom Scale; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PRO: Patientenberichteter Endpunkt.

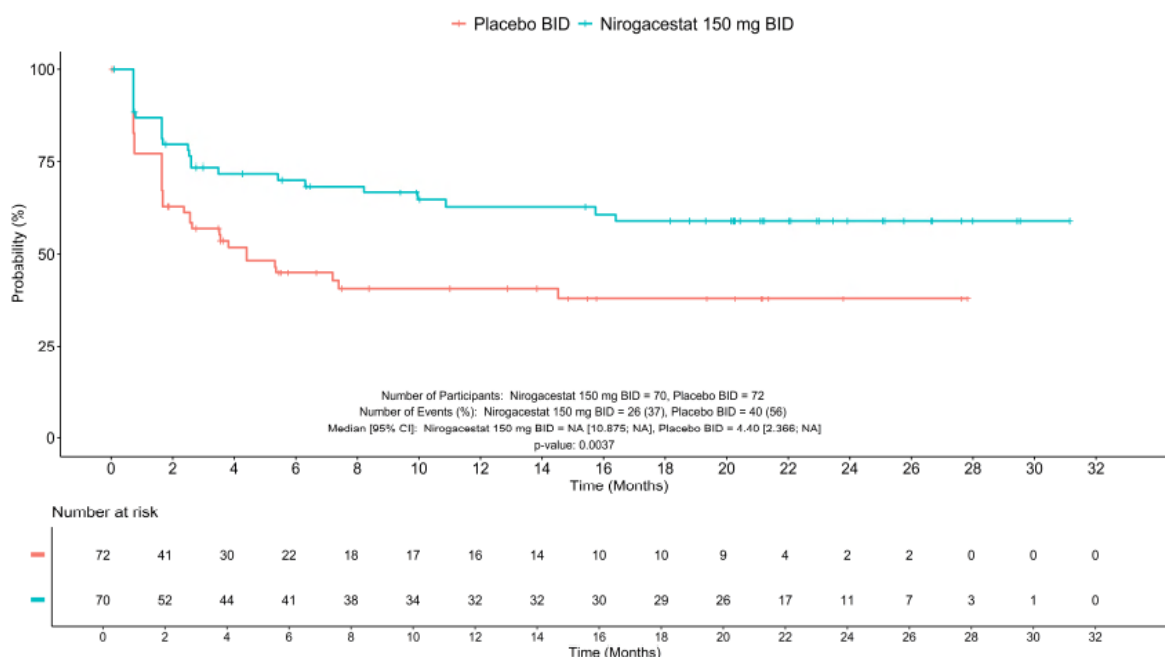


Abbildung 2: Kaplan-Meier-Kurve für GODDESS-DTSS – Gewichteter DTSS-Gesamtscore, Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

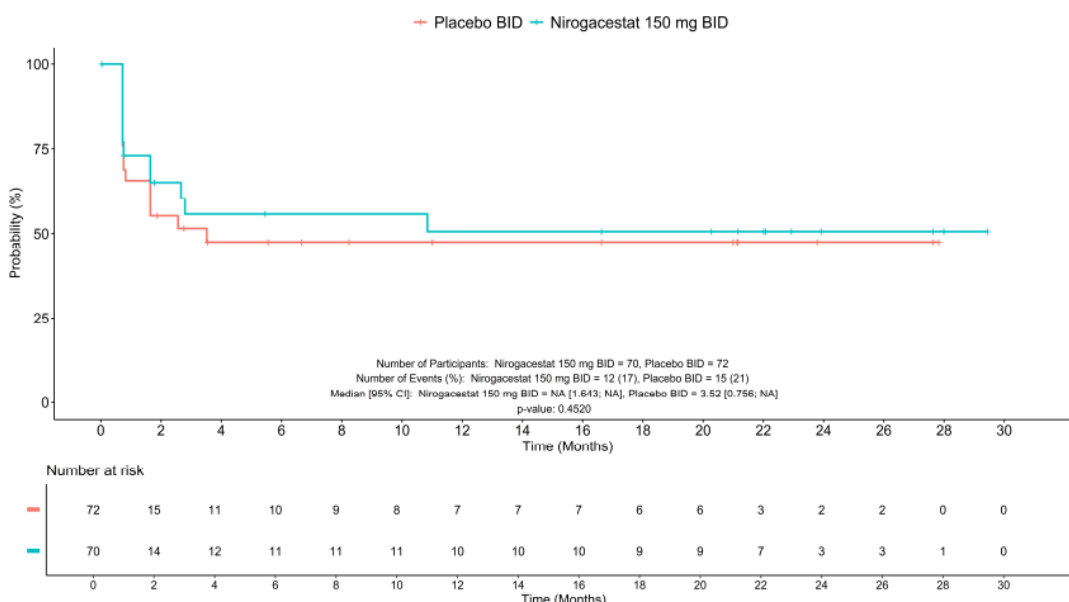


Abbildung 3: Kaplan-Meier-Kurve für GODDESS-DTSS – Domänenscore „Intra-abdominale Symptome“, Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

### Subgruppenanalysen

Für den gewichteten DTSS-Gesamtscore und den Domänenscore „Intra-abdominale Symptome“ liegen signifikante Interaktionsterme ( $p < 0,05$ ) vor. Die Effektschätzer in den jeweiligen Subgruppen waren für die vorliegende Nutzenbewertung nicht von Relevanz, bspw. aufgrund fehlender Signifikanz in den jeweiligen Subgruppen und/oder aufgrund fehlender klinischer Relevanz im Subgruppenmerkmal.

### **GODDESS-DTIS**

#### Responderanalyse

Für die GODDESS-DTIS werden in der vorliegenden Nutzenbewertung die 3 Domänenscores „Körperliche Funktion“, „Schlaf“ und „Emotion“ berücksichtigt. Auswertung zu den Einzelitems 9, 10 und 11 (Fragen zu „Körperwahrnehmung“, „Selbstbewusstsein“ und „Alltägliche praktische Herausforderungen“) wurden nicht vorgelegt.

Bei der Responderanalyse „Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite“ für die Domänenscores „Körperliche Funktion“ und „Schlaf“ zeigte sich ein signifikanter Unterschied zwischen den beiden Behandlungsarmen zugunsten von Nirogacestat (Tabelle 15, Abbildung 4 bis Abbildung 6).

*Tabelle 15: GODDESS-DTIS – Responderanalyse (Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)*

Studie DeFi GODDESS-DTIS <sup>1)</sup>	Nirogacestat N = 70	Placebo N = 72	HR [95%-KI] <sup>2)</sup> ; p-Wert <sup>3)</sup>
<b>Domänenscore „Körperliche Funktion“</b>			
Ereignisse (%)	16 (23)	35 (49)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	8,54 [6,18; 14,7]	0,33 [0,18; 0,60]; < 0,0001
<b>Domänenscore „Schlaf“</b>			
Ereignisse (%)	28 (40)	39 (54)	
Mediane Zeit (Monate) [95%-KI]	n. a. [8,28; n. a.]	7,36 [3,84; 11,1]	0,57 [0,34; 0,94] 0,01
<b>Domänenscore „Emotion“</b>			
Ereignisse (%)	30 (43)	37 (51)	
Mediane Zeit (Monate) [95%-KI]	n. a. [6,44; n. a.]	4,67 [2,83; 17,5]	0,61 [0,37; 1,01] 0,027

<sup>1)</sup> Eine Zensurierung erfolgte bei Personen, bei denen zum Zeitpunkt der Analyse keine Symptomverschlechterung beobachtet wurde. Diese wurden zum Zeitpunkt der letzten verfügbaren PRO-Bewertung zum oder vor dem Datenschnitt der Analysen zensiert.

<sup>2)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>3)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen.

Abkürzungen: DB: doppelblind; DTIS: Desmoid Tumor Impact Scale; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PRO: Patientenberichteter Endpunkt.

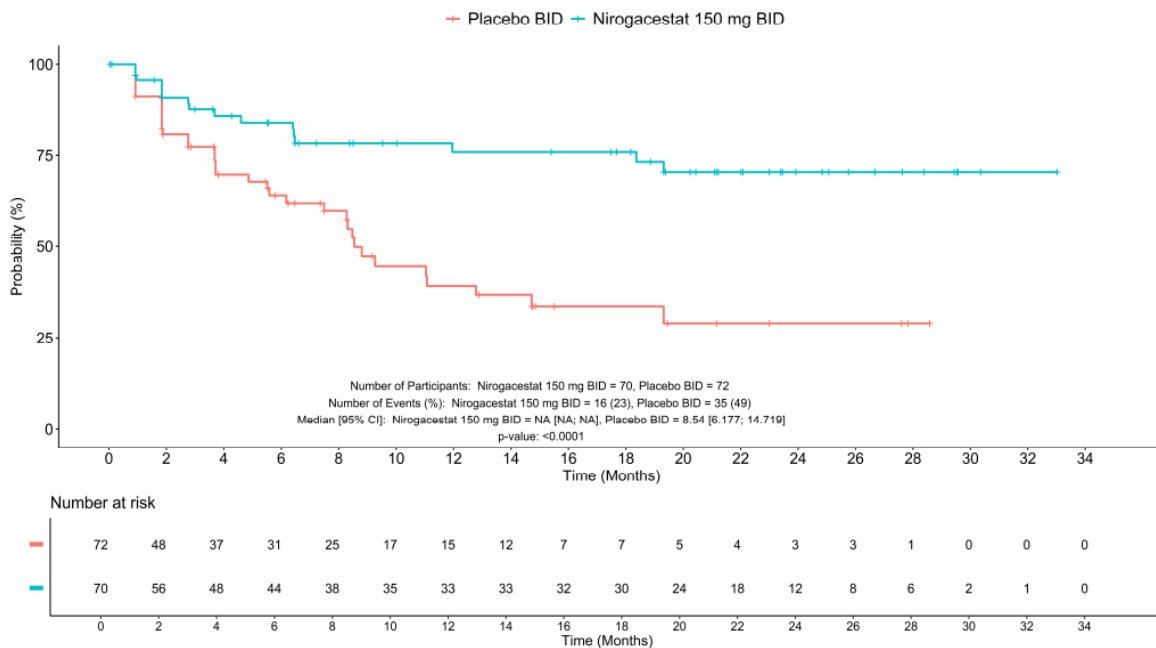


Abbildung 4: Kaplan-Meier-Kurve für GODDESS-DTIS – Domänenscore „Körperliche Funktion“, Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

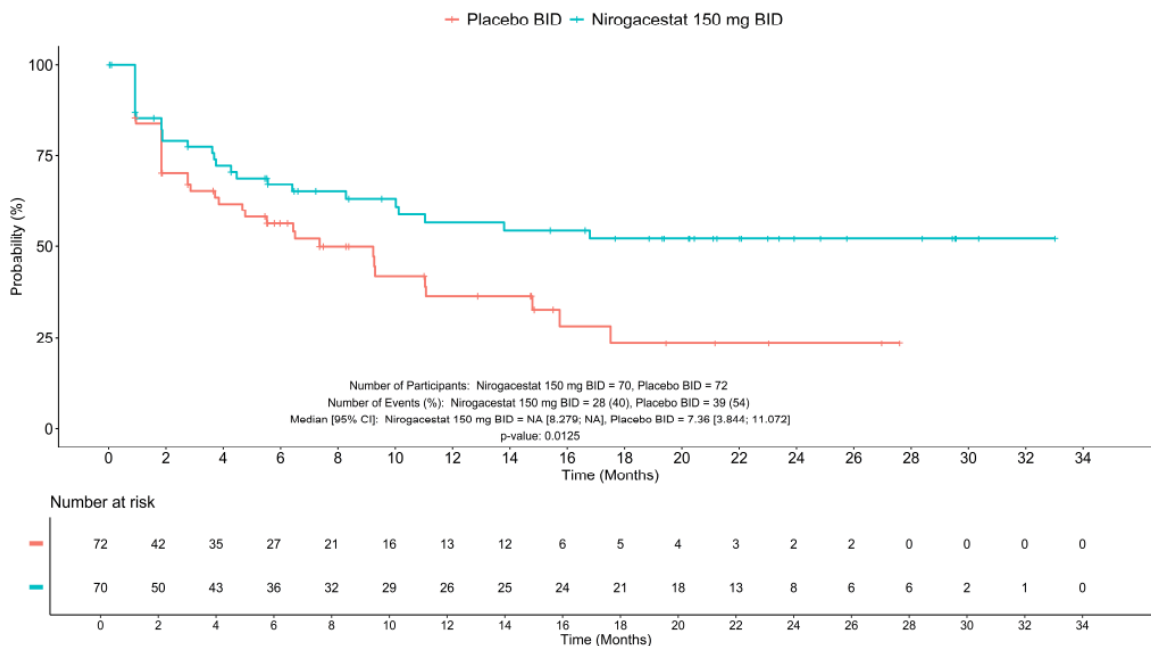


Abbildung 5: Kaplan-Meier-Kurve für GODDESS-DTIS – Domänenscore „Schlaf“, Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

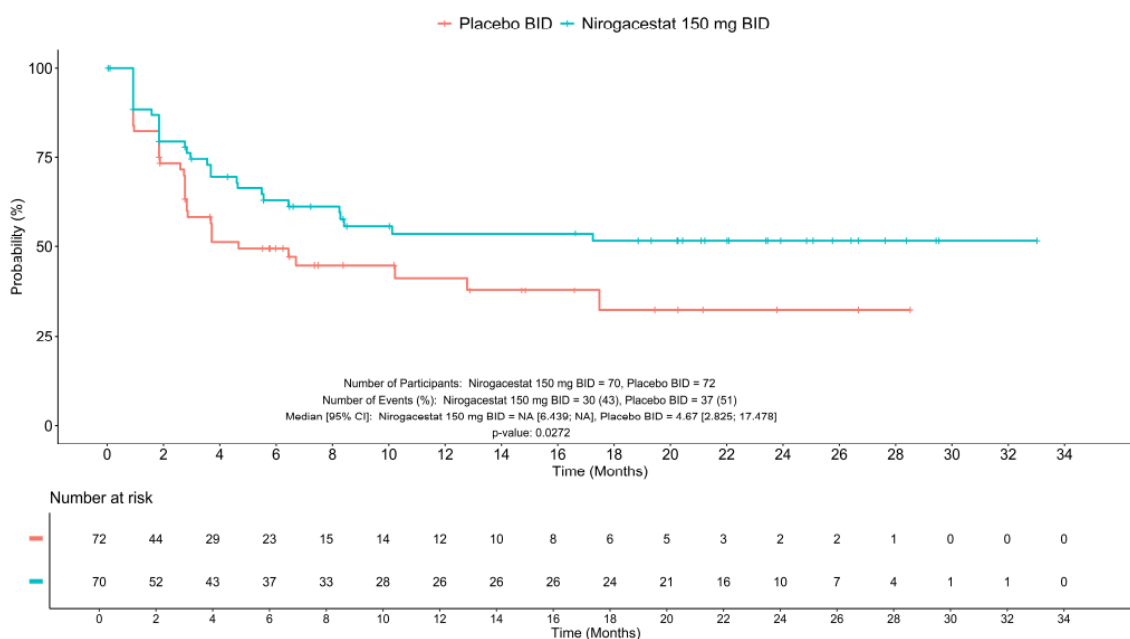


Abbildung 6: Kaplan-Meier-Kurve für GODDESS-DTIS – Domänenscore „Emotion“, Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

### Subgruppenanalysen

Für die GODDESS-DTIS lagen für die Domänenscores „Körperliche Funktion“ und „Schlaf“ signifikante ( $p < 0,05$ ) Interaktionsterme vor (siehe Tabelle 16). Für den Domänenscore „Schlaf“ lagen jedoch keine signifikanten relevanten Ergebnisse in den Subgruppen vor.

Tabelle 16: GODDESS-DTIS – Subgruppenanalysen (Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

Studie DeFi GODDESS-DTIS – Subgruppenanalysen		Nirogacestat N = 70	Placebo N = 72	Nirogacestat vs. Placebo HR [95%-KI] <sup>1)</sup> ; p-Wert <sup>2)</sup>
<b>Körperliche Funktion – Tumorlokalisation (<math>p = 0,0004</math>)<sup>3)</sup></b>				
Lokalisation der Baseline-Zielläsion (Stratifizierungsfaktor: extra-abdominal)	N	53 (100)	54 (100)	
	Ereignisse (%)	6 (11)	29 (54)	
	Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	8,48 [3,71;12,8]	0,13 [0,05; 0,32]; < 0,0001
Lokalisation der Baseline-Zielläsion (Stratifizierungsfaktor: intra-abdominal)	N	17 (100)	18 (100)	
	Ereignisse (%)	10 (59)	6 (33)	
	Mediane Zeit (Monate) [95%-KI]	12,0 [1,84; n. a.]	11,0 [6,18; n. a.]	1,58 [0,56; 4,48]; 0,19

Studie DeFi GODDESS-DTIS – Subgruppenanalysen		Nirogacestat N = 70	Placebo N = 72	Nirogacestat vs. Placebo HR [95%-KI] <sup>1)</sup> ; p-Wert <sup>2)</sup>
<b>Körperliche Funktion – Jegliche vorherige Therapie (p = 0,02)<sup>3)</sup></b>				
Jegliche vorherige Therapie (Ja)	N	52 (100)	58 (100)	
	Ereignisse (%)	7 (13)	29 (50)	
	Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	8,80 [7,49; 14,7]	0,17 [0,07; 0,41]; < 0,0001
Jegliche vorherige Therapie (Nein)	N	18 (100)	14 (100)	
	Ereignisse (%)	9 (50)	6 (43)	
	Mediane Zeit (Monate) [95%-KI]	12,0 [1,84; n. a.]	6,18 [1,84; n. a.]	0,72 [0,23; 2,25]; 0,29

<sup>1)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>2)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>3)</sup> P-Wert des Subgruppen-Interaktionstests ( $\alpha = 0,05$ ).

Abkürzungen: DB: doppelblind; DTIS: Desmoid Tumor Impact Scale; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar.

## Schmerzen mittels BPI-SF

### Responderanalyse

Für den BPI-SF werden in der vorliegenden Nutzenbewertung die zwei Skalen „Beeinträchtigung durch den Schmerz“ und „Schmerzintensität“ berücksichtigt (siehe Tabelle 17, Abbildung 7 und Abbildung 8).

Bei der Responderanalyse „Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite“ für die Subskala „Beeinträchtigung durch den Schmerz“ zeigte sich ein signifikanter Unterschied zwischen den beiden Behandlungsarmen zugunsten von Nirogacestat.

*Tabelle 17: BPI-SF – Responderanalyse (Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)*

Studie DeFi BPI-SF <sup>1)</sup>	Nirogacestat N = 70	Placebo N = 72	HR [95%-KI] <sup>2)</sup> ; p-Wert <sup>3)</sup>
<b>Skala „Beeinträchtigung durch den Schmerz“</b>			
Ereignisse (%)	12 (17)	22 (31)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	n. a. [n. a.; n. a.]	0,47 [0,23; 0,96] 0,02
<b>Skala „Schmerzintensität“</b>			
Ereignisse (%)	11 (16)	19 (26)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	n. a. [n. a.; n. a.]	0,50 [0,24; 1,06] 0,03

<sup>1)</sup> Bei Personen, bei denen zum Zeitpunkt der Analyse keine Symptomverschlechterung beobachtet wurde, erfolgte eine Zensurierung zum Zeitpunkt der letzten verfügbaren PRO-Bewertung zum oder vor dem Datenschnitt der Analyse.

2) Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

3) Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; DB: doppelblind; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PRO: Patientenberichteter Endpunkt.

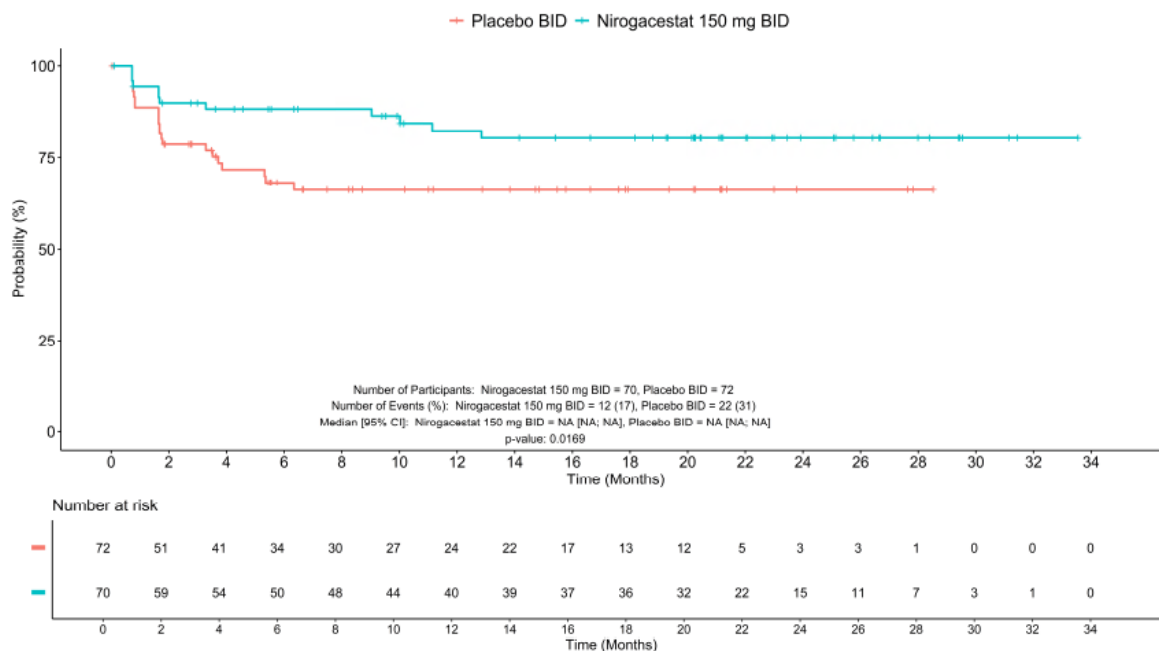
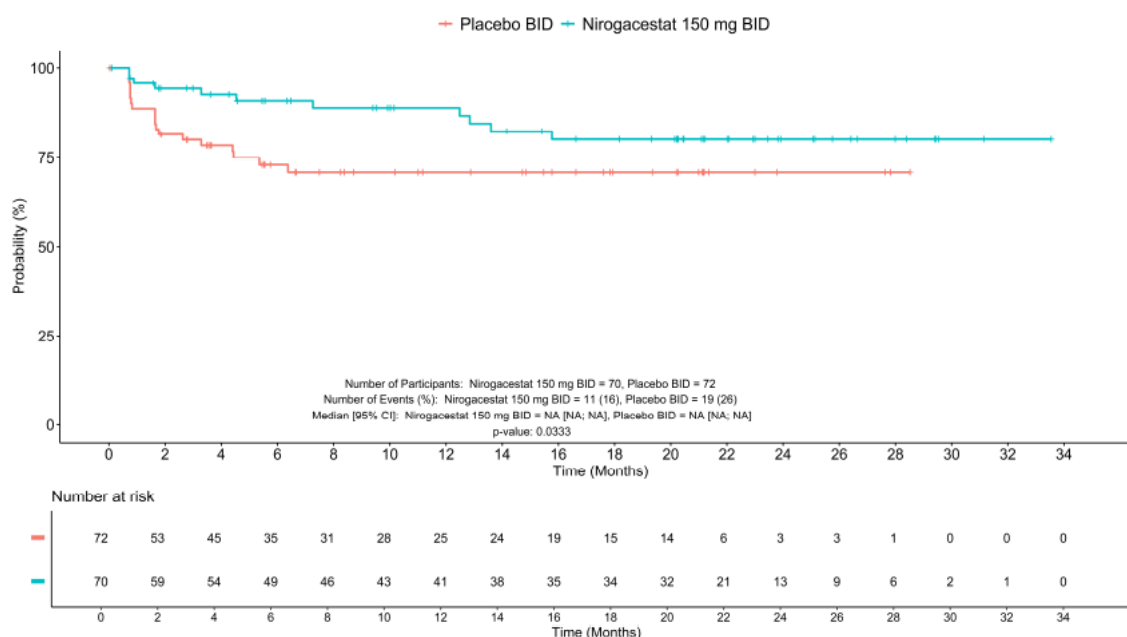


Abbildung 7: Kaplan-Meier-Kurve für BPI-SF – Skala „Beeinträchtigung durch den Schmerz“, Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)



**Abbildung 8:** Kaplan-Meier-Kurve für BPI-SF – Skala „Schmerzintensität“,  
 Zeit bis zur ersten Verschlechterung um  $\geq 15\%$  der Skalenspannweite;  
 Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

### Subgruppenanalysen

Für den BPI-SF lagen für die beide Skalen „Beeinträchtigung durch den Schmerz“ und „Schmerzintensität“ signifikante Interaktionsterme ( $p < 0,05$ ) vor. Die Effektschätzer in den jeweiligen Subgruppen waren für die vorliegende Nutzenbewertung nicht von Relevanz, bspw. aufgrund fehlender Signifikanz in den jeweiligen Subgruppen und/oder aufgrund fehlender klinischer Relevanz im Subgruppenmerkmal.

## Allgemeiner Gesundheitszustand mittels PGIC

### Responderanalyse

In Tabelle 18 werden die Ergebnisse der Responderanalyse (Zeit bis zu jeglicher erstmaligen Verschlechterung) dargestellt. In Abbildung 9 zeigt sich, dass sich der therapeutische Nutzen überwiegend im späteren Verlauf manifestiert, während in der frühen Phase kein klarer Vorteil erkennbar ist. Die Verletzung der Proportional-Hazard-Annahme war statistisch signifikant ( $p = 0,002$ ).

Tabelle 18: PGIC – Responderanalyse (Zeit bis zu jeglicher erstmaligen Verschlechterung); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

Studie DeFi PGIC	Nirogacestat N = 70	Placebo N = 72	HR [95%-KI] <sup>1</sup> ; p-Wert <sup>2</sup>
<b>Zeit bis zu jeglicher erstmaligen Verschlechterung im PGIC</b>			
Ereignisse (%)	32 (46)	47 (65)	
Mediane Zeit (Monate) [95%-KI]	22,3 [3,75; n. a.]	4,86 [3,48; 8,28]	0,53 <sup>3</sup> [0,33; 0,86]; 0,004

<sup>1</sup>) Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>2</sup>) Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>3</sup>) Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,002$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Abkürzungen: DB: doppelblind; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PGIC: Patient Global Impression of Change;

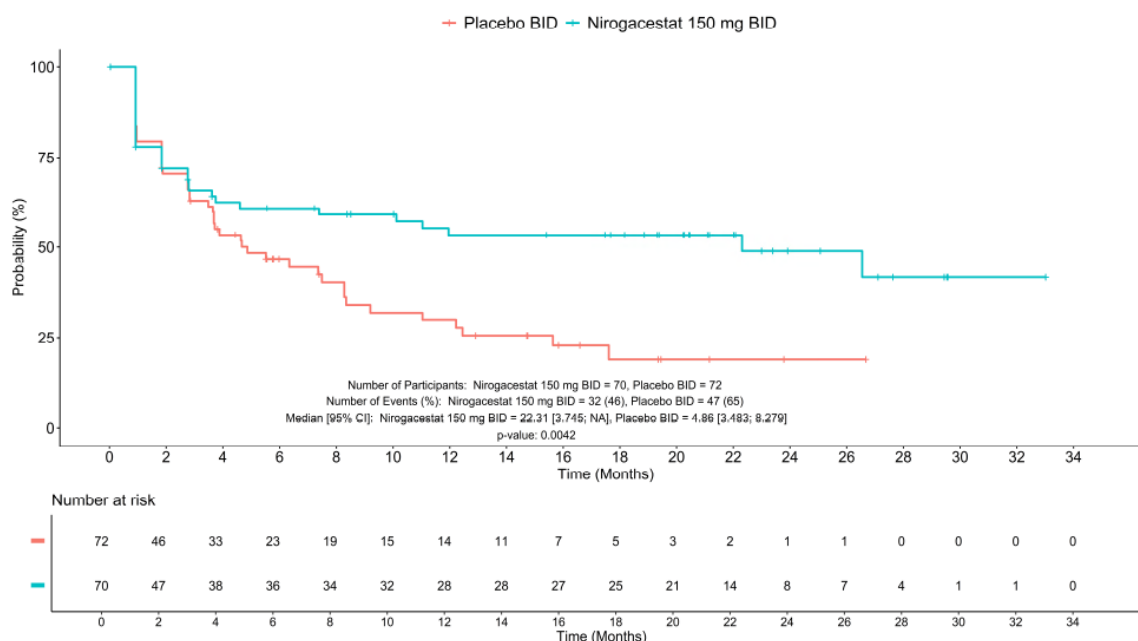


Abbildung 9: Kaplan-Meier-Kurve für den PGIC (allgemeiner Gesundheitszustand), Zeit bis zu jeglicher erstmaligen Verschlechterung, Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

### Subgruppenanalysen

Es lagen signifikante ( $p < 0,05$ ) Interaktionsterme für den PGIC (Zeit bis zu jeglicher Verschlechterung) vor. Für die vorliegende Nutzenbewertung relevante Subgruppeneffekte, mit mindestens einem signifikanten Effektschätzer in den jeweiligen Subgruppen, werden in Tabelle 19 dargestellt.

*Tabelle 19: PGIC – Subgruppenanalyse (Zeit bis zu jeglicher Verschlechterung; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022))*

Studie DeFi PGIC – Subgruppenanalyse		Nirogacestat N = 70	Placebo N = 72	Nirogacestat vs. Placebo HR [95%-KI] <sup>1)</sup> ; p-Wert <sup>2)</sup>
<b>PGIC– Alter (&lt; 35 Jahre; ≥ 35 Jahre) (p = 0,02)<sup>3</sup></b>				
Alter < 35 Jahre	N	38 (100)	36 (100)	
	Ereignisse (%)	11 (29)	22 (61)	
	Mediane Zeit (Monate) [95%-KI]	n. a. [22,3; n. a.]	6,34 [1,87; 9,20]	0,28 [0,13; 0,60]; 0,0003
Alter ≥ 35 Jahre	N	32 (100)	36 (100)	
	Ereignisse (%)	21 (66)	25 (69)	
	Mediane Zeit (Monate) [95%-KI]	2,79 [0,92; 10,1]	4,67 [2,83; 11,0]	1,07 <sup>4)</sup> [0,57; 2,00]; 0,42

<sup>1)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>2)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>3)</sup> P-Wert des Subgruppen-Interaktionstests ( $\alpha = 0,05$ ).

<sup>4)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,03$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Abkürzungen: DB: doppelblind; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PGIC: Patient Global Impression of Change.

## Symptomatik mittels EORTC QLQ-C30: Symptomskalen

### Responderanalyse

Die Responderanalysen „Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte“ im Rahmen der Symptomskalen des EORTC QLQ-C30 sind Tabelle 20 zu entnehmen. Im Anhang finden sich die Kaplan-Meier-Kurven für die entsprechenden Symptomskalen des EORTC QLQ-C30 (Abbildung 10 bis Abbildung 17).

Bei der Responderanalyse „Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte“ zeigte sich für die Symptomskalen „Übelkeit und Erbrechen“, „Appetitverlust“ und „Diarrhö“ ein signifikanter Unterschied zwischen den beiden Behandlungsgruppen zuungunsten von Nirogacestat. Für die Symptomskala „Obstipation“ zeigte sich ein signifikanter Unterschied zwischen den beiden Behandlungsgruppen zugunsten von Nirogacestat.

*Tabelle 20: EORTC QLQ-C30: Symptomskalen – Responderanalyse (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)*

Studie DeFi EORTC QLQ-C30: Symptomskalen <sup>1)</sup>	Nirogacestat N = 70	Placebo N = 72	HR [95%-KI] <sup>2)</sup> ; p-Wert <sup>3)</sup>
<b>Fatigue</b>			
Ereignisse (%)	48 (69)	45 (63)	
Mediane Zeit (Monate) [95%-KI]	1,84 [0,95; 4,63]	2,76 [1,84; 5,55]	1,05 [0,66; 1,65]; 0,42
<b>Übelkeit und Erbrechen</b>			
Ereignisse (%)	51 (73)	28 (39)	
Mediane Zeit (Monate) [95%-KI]	0,92 [0,92; 2,73]	21,0 [9,20; n. a.]	3,00 [1,80; 5,00]; < 0,0001 <sup>4)</sup>
<b>Schmerz</b>			
Ereignisse (%)	28 (40)	36 (50)	
Mediane Zeit (Monate) [95%-KI]	n. a. [7,49; n. a.]	6,44 [3,48; n. a.]	0,62 [0,37; 1,05]; 0,04
<b>Dyspnoe</b>			
Ereignisse (%)	31 (44)	27 (38)	
Mediane Zeit (Monate) [95%-KI]	19,3 [7,33; n. a.]	11,0 [6,57; n. a.]	1,07 [0,63; 1,82]; 0,40
<b>Schlaflosigkeit</b>			
Ereignisse (%)	28 (40)	28 (39)	
Mediane Zeit (Monate) [95%-KI]	23,0 [6,44; n. a.]	18,4 [6,44; n. a.]	0,88 [0,52; 1,52]; 0,33
<b>Appetitverlust</b>			
Ereignisse (%)	37 (53)	23 (32)	
Mediane Zeit (Monate) [95%-KI]	6,44 [1,84; n. a.]	n. a. [8,54; n. a.]	1,96 [1,13; 3,41]; 0,01

Studie DeFi EORTC QLQ-C30: Symptomskalen <sup>1)</sup>	Nirogacestat N = 70	Placebo N = 72	HR [95%-KI] <sup>2)</sup> ; p-Wert <sup>3)</sup>
<b>Obstipation</b>			
Ereignisse (%)	16 (23)	27 (38)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	n. a. [4,40; n. a.]	0,49 [0,26; 0,92]; 0,01
<b>Diarrhö</b>			
Ereignisse (%)	55 (79)	32 (44)	
Mediane Zeit (Monate) [95%-KI]	0,92 [0,92; 1,84]	9,23 [5,72; n. a.]	3,58 [2,20; 5,84]; < 0,0001

<sup>1)</sup> Eine Zensurierung erfolgte bei Personen, bei denen zum Zeitpunkt der Analyse keine Symptomverschlechterung beobachtet wurde. Diese wurden zum Zeitpunkt der letzten verfügbaren PRO-Bewertung zum oder vor dem Datenschnitt der Analysen zensiert.

<sup>2)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>3)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>4)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,03$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Abkürzungen: DB: doppelblind; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PRO: Patientenberichteter Endpunkt.

### Subgruppenanalysen

Für den EORTC-QLQ C30 lagen für die Symptomskalen signifikante Interaktionsterme ( $p < 0,05$ ) vor. Die Effektschätzer in den jeweiligen Subgruppen waren für die vorliegende Nutzenbewertung nicht von Relevanz, bspw. aufgrund fehlender Signifikanz in den jeweiligen Subgruppen und/oder aufgrund fehlender klinischer Relevanz im Subgruppenmerkmal.

## 3.4 Lebensqualität

### EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus

#### Responderanalyse

Die Responderanalysen „Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte“ im Rahmen der Funktionsskalen und der globalen Skala „Allgemeiner Gesundheitsstatus / Lebensqualität“ des EORTC QLQ-C30 sind Tabelle 21 zu entnehmen. Im Anhang finden sich die Kaplan-Meier-Kurven für die entsprechenden Funktionsskalen des EORTC QLQ-C30 (Abbildung 18 bis Abbildung 23).

Bei der Responderanalyse „Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte“ zeigte sich für die Funktionsskala „Körperliche Funktion“ ein signifikanter Unterschied zwischen den beiden Behandlungsgruppen zugunsten von Nirogacestat.

**Tabelle 21: EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus –Responderanalysen (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)**

<b>Studie DeFi EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus<sup>1)</sup></b>	<b>Nirogacestat N = 70</b>	<b>Placebo N = 72</b>	<b>HR [95%-KI]<sup>2)</sup>; p-Wert<sup>3)</sup></b>
<b>Globale Skala „Allgemeiner Gesundheitsstatus / Lebensqualität“</b>			
Ereignisse (%)	41 (59)	40 (56)	
Mediane Zeit (Monate) [95%-KI]	6,21 [2,83; 18,9]	6,47 [3,71; 11,0]	0,95 [0,60; 1,50]; 0,41
<b>Kognitive Funktion</b>			
Ereignisse (%)	40 (57)	32 (44)	
Mediane Zeit (Monate) [95%-KI]	4,86 [2,79; 17,5]	12,0 [5,42; 16,6]	1,35 [0,83; 2,19]; 0,11
<b>Emotionale Funktion</b>			
Ereignisse (%)	34 (49)	36 (50)	
Mediane Zeit (Monate) [95%-KI]	13,8 [6,44; n. a.]	8,28 [3,90; 16,5]	0,77 [0,48; 1,25]; 0,15
<b>Körperliche Funktion</b>			
Ereignisse (%)	16 (23)	33 (46)	
Mediane Zeit (Monate) [95%-KI]	n. a. [n. a.; n. a.]	8,38 [7,39; 18,8]	0,38 [0,21; 0,70]; 0,001
<b>Rollenfunktion</b>			
Ereignisse (%)	32 (46)	37 (51)	
Mediane Zeit (Monate) [95%-KI]	19,3 [4,67; n. a.]	7,56 [4,67; 20,3]	0,75 [0,46; 1,23]; 0,12
<b>Soziale Funktion</b>			
Ereignisse (%)	38 (54)	39 (54)	
Mediane Zeit (Monate) [95%-KI]	6,44 [3,45; n. a.]	5,55 [3,52; 8,54]	0,87 [0,54; 1,40]; 0,28

<sup>1)</sup> Eine Zensurierung erfolgte bei Personen, bei denen zum Zeitpunkt der Analyse keine Symptomverschlechterung beobachtet wurde. Diese wurden zum Zeitpunkt der letzten verfügbaren PRO-Bewertung zum oder vor dem Datenschnitt der Analysen zensiert.

<sup>2)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>3)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen.

Abkürzungen: DB: doppelblind; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; HR: Hazard Ratio; ITT: Intention to Treat; KI: Konfidenzintervall; n. a.: nicht anwendbar; PRO: Patientenberichteter Endpunkt.

### Subgruppenanalysen

Es lagen keine signifikanten ( $p < 0,05$ ) Interaktionsterme für die Funktionsskalen und die globale Skala „Allgemeiner Gesundheitsstatus / Lebensqualität“ des EORTC QLQ-C30 vor.

### 3.5 Sicherheit

Für schwere UE (CTCAE-Grad  $\geq 3$ ) sowie UE, die zum Abbruch der Studienmedikation führten, zeigte sich ein statistisch signifikanter Behandlungsunterschied zuungunsten von Nirogacestat gegenüber Placebo. Die zugehörigen Kaplan-Meier-Kurven sind dem Anhang (Abbildung 24 bis Abbildung 26) zu entnehmen.

Ereignisse wurden nicht als UE gewertet, wenn sie im Zusammenhang mit der Grunderkrankung standen. Allerdings wurden hierfür keine Kriterien definiert. Basierend darauf kann nicht abschließend bewertet werden, ob Ereignisse der Grunderkrankung mit in die Erfassung der UE eingingen.

Die mediane Behandlungsdauer betrug im Nirogacestat-Arm 20,6 Monate (min; max: 0,3; 33,6) und lag somit über der medianen Behandlungsdauer im Placebo-Arm (11,4 Monate (min; max: 0,2; 32,5)). Angaben zur Beobachtungsdauer für UE liegen nicht vor.

*Tabelle 22: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022)*

<b>Studie DeFi Zusammenfassung der UE<sup>1)</sup></b> <i>Personen mit mind. einem ...</i>	<b>Nirogacestat<sup>2)</sup></b> <b>N = 69</b> <i>Mediane Zeit bis Ereignis, Monate [95%-KI] Ereignis, n (%)</i>	<b>Placebo<sup>2)</sup></b> <b>N = 72</b> <i>Mediane Zeit bis Ereignis, Monate [95%-KI] Ereignis, n (%)</i>	<b>HR<sup>3)</sup> [95%-KI]; p-Wert<sup>4)</sup></b>
UE (ergänzend dargestellt)	0,10 [0,03; 0,16] 69 (100)	0,26 [0,10; 0,30] 69 (96)	-
schweren UE (CTCAE-Grad $\geq 3$ )	17,5 [3,32;27,6] 38 (55)	n. a. [21,3; n. a.] 12 (17)	3,80 [1,98; 7,31]; < 0,0001
SUE	n. a. [n. a.; n. a.] 14 (20)	n. a. [21,3; n. a.] 8 (11)	1,77 <sup>5)</sup> [0,74; 4,26]; 0,10
UE, das zum Abbruch der Studienmedikation führte	n. a. [n. a.; n. a.] 16 (23)	n. a. [22,7; n. a.] 2 (3)	8,26 <sup>5)</sup> [1,90;36,0]; 0,0004

<sup>1)</sup> Definiert als jedes unerwünschte medizinische Ereignis mit anfänglichem Auftreten oder zunehmendem Schweregrad nach Einnahme der ersten Dosis der Studienbehandlung bis 30 Tage nach der letzten Dosis der Studienmedikation. Wenn eine Person ein kodiertes Sicherheitsereignis mehr als einmal erlebte, wurde dieses Ereignis nur einmal gezählt.

<sup>2)</sup> Die Sicherheitspopulation umfasst alle Personen, welche zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben.

<sup>3)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>4)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>5)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,02$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; DB: doppelblind; HR: Hazard Ratio; KI: Konfidenzintervall; n. a.: nicht anwendbar; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## Unerwünschte Ereignisse

Für die folgenden SOC zeigten sich in den Ereigniszeitanalysen statistisch signifikante Unterschiede zuungunsten von Nirogacestat: „Erkrankungen des Gastrointestinaltrakts“, „Infektionen und parasitäre Erkrankungen“, „Stoffwechsel- und Ernährungsstörungen“, „Erkrankungen der Geschlechtsorgane und der Brustdrüse“, „Erkrankungen der Atemwege, des Brustraums und Mediastinums“, „Erkrankungen der Haut und des Unterhautgewebes“ sowie „Gefäßerkrankungen“. In der SOC „Gutartige, bösartige und unspezifische Neubildungen (einschl. Zysten und Polypen)“ zeigten sich statistisch signifikante Unterschiede zugunsten von Nirogacestat.

*Tabelle 23: UE mit Inzidenz  $\geq 10\%$  in mindestens einem Studienarm nach Systemorganklasse und Preferred Term, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022)*

Studie DeFi UE <sup>1)</sup> <i>MedDRA-Systemorganklasse<sup>2)</sup> Preferred Term</i>	Nirogacestat <sup>3)</sup> N = 69 <i>Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</i>	Placebo <sup>3)</sup> N = 72 <i>Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</i>	HR <sup>4)</sup> [95%-KI]; p-Wert <sup>5)</sup>
<b>Erkrankungen des Blutes und des Lymphsystems</b>	<b>n. a. [n. a.; n. a.] 11 (16)</b>	<b>n. a. [n. a.; n. a.] 9 (13)</b>	<b>1,26 [0,52; 3,05]; 0,30</b>
Anämie	n. a. [n. a.; n. a.] 7 (10)	n. a. [n. a.; n. a.] 8 (11)	0,89 [0,32; 2,46]; 0,41
<b>Herzerkrankungen</b>	<b>n. a. [n. a.; n. a.] 9 (13)</b>	<b>n. a. [n. a.; n. a.] 3 (4)</b>	<b>2,99 [0,80; 11,1]; 0,04</b>
<b>Augenerkrankungen</b>	<b>n. a. [n. a.; n. a.] 8 (12)</b>	<b>n. a. [n. a.; n. a.] 4 (6)</b>	<b>2,08 [0,62; 6,92]; 0,11</b>
<b>Erkrankungen des Gastrointestinaltrakts</b>	<b>0,26 [0,16; 0,26] 67 (97)</b>	<b>2,60 [0,53; 7,23] 47 (65)</b>	<b>3,43 [2,27; 5,16]; &lt; 0,0001</b>
Abdominalschmerz	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 9 (13)	1,15 [0,47; 2,80]; 0,38
Obstipation	n. a. [n. a.; n. a.] 7 (10)	n. a. [n. a.; n. a.] 7 (10)	0,96 [0,33; 2,74]; 0,47
Diarrhö	0,33 [0,26; 0,66] 58 (84)	n. a. [12,2; n. a.] 25 (35)	4,30 [2,66; 6,95]; < 0,0001
Mundtrockenheit	n. a. [n. a.; n. a.] 8 (12)	n. a. [n. a.; n. a.] 3 (4)	2,84 [0,75; 10,7]; 0,05
Übelkeit	8,31 [0,43; n. a.] 37 (54)	22,9 [14,0; n. a.] 28 (39)	1,69 <sup>6)</sup> [1,03; 2,77]; 0,02
Stomatitis	n. a. [n. a.; n. a.] 20 (29)	n. a. [n. a.; n. a.] 3 (4)	7,94 [2,35; 26,8]; < 0,0001
Erbrechen	n. a. [n. a.; n. a.] 14 (20)	n. a. [22,0; n. a.] 14 (19)	1,01 [0,48; 2,13]; 0,49
<b>Allgemeine Erkrankungen und Beschwerden am Verabreichungsort</b>	<b>2,76 [0,76; 13,3] 44 (64)</b>	<b>13,8 [2,79; n. a.] 36 (50)</b>	<b>1,44 [0,92; 2,25]; 0,05</b>
Ermüdung	21,6 [0,95; n. a.] 35 (51)	n. a. [22,7; n. a.] 26 (36)	1,52 [0,91; 2,55]; 0,05
Grippeähnliche Erkrankung	n. a. [n. a.; n. a.] 7 (10)	n. a. [n. a.; n. a.] 2 (3)	3,21 [0,66; 15,5]; 0,06

<b>Studie DeFi UE<sup>1)</sup> MedDRA-Systemorganklasse<sup>2)</sup> Preferred Term</b>	<b>Nirogacestat<sup>3)</sup> N = 69 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>Placebo<sup>3)</sup> N = 72 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>HR<sup>4)</sup> [95%-KI]; p-Wert<sup>5)</sup></b>
<b>Infektionen und parasitäre Erkrankungen</b>	<b>11,1 [5,39; 21,8] 39 (57)</b>	<b>n. a. [17,0; n. a.] 24 (33)</b>	<b>1,90 [1,14; 3,17]; 0,01</b>
COVID-19	30,5 [26,3; n. a.] 12 (17)	n. a. [n. a.; n. a.] 12 (17)	0,66 <sup>6)</sup> [0,29; 1,49]; 0,16
Follikulitis <sup>10)</sup>	n. a. [n. a.; n. a.] 9 (13)	n. a. [n. a.; n. a.] 0 (0)	- <sup>7)</sup>
<b>Verletzung, Vergiftung u. durch Eingriffe bedingte Komplikationen</b>	<b>n. a. [n. a.; n. a.] 9 (13)</b>	<b>n. a. [n. a.; n. a.] 5 (7)</b>	<b>1,50<sup>8)</sup> [0,49; 4,56]; 0,24</b>
<b>Untersuchungen</b>	<b>11,1 [2,66; 19,1] 40 (58)</b>	<b>17,0 [7,69; n. a.] 33 (46)</b>	<b>1,37<sup>9)</sup> [0,85; 2,20]; 0,10</b>
Alanin-Aminotransferase erhöht <sup>10)</sup>	n. a. [n. a.; n. a.] 12 (17)	n. a. [n. a.; n. a.] 6 (8)	2,11 [0,79; 5,66]; 0,06
Aspartat-Aminotransferase erhöht <sup>10)</sup>	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 8 (11)	1,41 [0,56; 3,54]; 0,23
Follikelstimulierendes Hormon im Blut erhöht	n. a. [n. a.; n. a.] 7 (10)	n. a. [n. a.; n. a.] 1 (1)	7,46 [0,92; 60,7]; 0,01
SARS-CoV-2-Test positiv	n. a. [26,3; n. a.] 8 (12)	n. a. [n. a.; n. a.] 10 (14)	0,56 [0,22; 1,45]; 0,11
Gewicht erhöht	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 5 (7)	2,02 [0,70; 5,84]; 0,09
<b>Stoffwechsel- und Ernährungsstörungen</b>	<b>2,79 [0,53; 11,1] 43 (62)</b>	<b>n. a. [n. a.; n. a.] 21 (29)</b>	<b>2,90 [1,70; 4,94] &lt; 0,0001</b>
Appetit vermindert	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 8 (11)	1,47 [0,59; 3,66]; 0,20
Hypokaliämie	n. a. [n. a.; n. a.] 8 (12)	n. a. [n. a.; n. a.] 1 (1)	8,90 [1,11; 71,4] 0,01
Hypophosphatämie <sup>10)</sup>	27,4 [8,31; n. a.] 29 (42)	n. a. [n. a.; n. a.] 5 (7)	7,36 [2,83; 19,1]; < 0,0001
<b>Skelettmuskulatur-, Bindege- webs- u. Knochenkrankungen</b>	<b>21,5 [13,3; n. a.] 32 (46)</b>	<b>23,5 [10,4; n. a.] 30 (42)</b>	<b>0,96 [0,58; 1,59]; 0,43</b>
Arthralgie	30,8 [30,8; n. a.] 8 (12)	n. a. [23,5; n. a.] 9 (13)	0,62 [0,23; 1,68] 0,17
Rückenschmerzen	n. a. [n. a.; n. a.] 9 (13)	n. a. [n. a.; n. a.] 7 (10)	1,11 [0,41; 3,00]; 0,42
<b>Gutartige, bösartige und unspezifische Neubildungen (einschl. Zysten und Polypen)</b>	<b>n. a. [n. a.; n. a.] 9 (13)</b>	<b>n. a. [n. a.; n. a.] 17 (24)</b>	<b>0,43<sup>11)</sup> [0,19; 0,96]; 0,02</b>
Tumorschmerzen	n. a. [n. a.; n. a.] 5 (7)	n. a. [n. a.; n. a.] 13 (18)	0,33 <sup>12)</sup> [0,12; 0,93]; 0,01

<b>Studie DeFi UE<sup>1)</sup> MedDRA-Systemorganklasse<sup>2)</sup> Preferred Term</b>	<b>Nirogacestat<sup>3)</sup> N = 69 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>Placebo<sup>3)</sup> N = 72 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>HR<sup>4)</sup> [95%-KI]; p-Wert<sup>5)</sup></b>
<b>Erkrankungen des Nervensystems</b>	<b>13,3 [7,59; n. a.] 34 (49)</b>	<b>n. a. [18,6; n. a.] 26 (36)</b>	<b>1,46 [0,87; 2,43]; 0,07</b>
Schwindelgefühl	n. a. [n. a.; n. a.] 8 (12)	n. a. [n. a.; n. a.] 4 (6)	1,83 [0,55; 6,15]; 0,16
Kopfschmerzen	n. a. [26,7; n. a.] 20 (29)	n. a. [n. a.; n. a.] 11 (15)	1,99 [0,95; 4,18]; 0,03
<b>Psychiatrische Erkrankungen</b>	<b>27,7 [27,1; n. a.] 17 (25)</b>	<b>n. a. [22,7; n. a.] 12 (17)</b>	<b>1,47 [0,70; 3,10]; 0,15</b>
Einschlafstörungen	n. a. [n. a.; n. a.] 6 (9)	n. a. [n. a.; n. a.] 8 (11)	0,75 [0,26; 2,16]; 0,29
<b>Erkrankungen der Nieren und Harnwege</b>	<b>n. a. [n. a.; n. a.] 9 (13)</b>	<b>n. a. [n. a.; n. a.] 7 (10)</b>	<b>1,25 [0,46; 3,37]; 0,33</b>
<b>Erkrankungen der Geschlechts- organe und der Brustdrüse</b>	<b>25,1 [2,99; n. a.] 32 (46)</b>	<b>n. a. [n. a.; n. a.] 12 (17)</b>	<b>3,42<sup>12)</sup> [1,76; 6,68]; &lt; 0,0001</b>
Nachlassen der ovariellen Funktion	n. a. [n. a.; n. a.] 13 (19)	n. a. [n. a.; n. a.] 0 (0)	<sup>7)</sup>
Vorzeitige Menopause <sup>10)</sup>	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 0 (0)	<sup>7)</sup>
<b>Erkrankungen der Atemwege, des Brustraums u. Mediastinums</b>	<b>n. a. [6,34; n. a.] 29 (42)</b>	<b>n. a. [n. a.; n. a.] 14 (19)</b>	<b>2,51 [1,33; 4,77]; 0,002</b>
Husten	n. a. [n. a.; n. a.] 11 (16)	n. a. [22,7; n. a.] 3 (4)	3,78 [1,05; 13,6]; 0,01
Dyspnoe	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 4 (6)	2,82 [0,89; 8,89]; 0,03
Epistaxis	n. a. [n. a.; n. a.] 10 (14)	n. a. [n. a.; n. a.] 1 (1)	11,3 [1,45; 88,6]; 0,002
Verstopfte Nase	n. a. [n. a.; n. a.] 7 (10)	n. a. [n. a.; n. a.] 2 (3)	3,71 [0,77; 17,9]; 0,04
<b>Erkrankungen der Haut und des Unterhautgewebes</b>	<b>1,48 [0,72; 2,96] 55 (80)</b>	<b>n. a. [14,9; n. a.] 23 (32)</b>	<b>3,91 [2,38; 6,40]; &lt; 0,0001</b>
Alopezie	n. a. [n. a.; n. a.] 13 (19)	n. a. [n. a.; n. a.] 1 (1)	11,5 [1,49; 88,2]; 0,002
Dermatitis akneiform	n. a. [n. a.; n. a.] 15 (22)	n. a. [n. a.; n. a.] 0 (0)	<sup>7)</sup>
Trockene Haut	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 5 (7)	2,12 [0,74; 6,13]; 0,08
Pruritus	n. a. [n. a.; n. a.] 9 (13)	n. a. [n. a.; n. a.] 6 (8)	1,49 [0,53; 4,19]; 0,22
Ausschlag	n. a. [n. a.; n. a.] 13 (19)	n. a. [n. a.; n. a.] 5 (7)	2,49 [0,88; 7,03]; 0,04
Makulopapulöser Ausschlag <sup>10)</sup>	n. a. [11,9; n. a.] 22 (32)	n. a. [n. a.; n. a.] 4 (6)	6,37 [2,19; 18,5]; < 0,0001

<b>Studie DeFi UE<sup>1)</sup> MedDRA-Systemorganklasse<sup>2)</sup> Preferred Term</b>	<b>Nirogacestat<sup>3)</sup> N = 69 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>Placebo<sup>3)</sup> N = 72 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>HR<sup>4)</sup> [95%-KI]; p-Wert<sup>5)</sup></b>
<b>Gefäßerkrankungen</b>	<b>n. a. [11,1; n. a.] 26 (38)</b>	<b>n. a. [22,1; n. a.] 14 (19)</b>	<b>2,00 [1,04; 3,86]; 0,02</b>
Hitzewallungen	n. a. [n. a.; n. a.] 13 (19)	n. a. [n. a.; n. a.] 4 (6)	3,41 [1,11; 10,5]; 0,01
Hypertonie	n. a. [n. a.; n. a.] 10 (14)	n. a. [n. a.; n. a.] 8 (11)	1,23 [0,48; 3,15]; 0,33

<sup>1)</sup> Definiert als jedes unerwünschte medizinische Ereignis mit anfänglichem Auftreten oder zunehmenden Schweregrad nach Einnahme der ersten Dosis der Studienbehandlung bis 30 Tage nach der letzten Dosis der Studienmedikation. Wenn eine Person ein kodiertes Sicherheitsereignis mehr als einmal erlebte, wurde dieses Ereignis nur einmal gezählt.

<sup>2)</sup> Die Kodierung erfolgte nach MedDRA-Version 24.0.

<sup>3)</sup> Die Sicherheitspopulation umfasst alle Personen, die zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben.

<sup>4)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>5)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>6)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,01$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>7)</sup> Die Berechnung eines HR konnte aufgrund fehlender Ereignisse im Placebo-Arm nicht erfolgen und Log-Rank-Test wurde nicht berichtet.

<sup>8)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,045$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>9)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,04$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>10)</sup> Definiert als UESI.

<sup>11)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,001$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>12)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,02$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Abkürzungen: DB: doppelblind; HR: Hazard Ratio; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; n. a.: nicht anwendbar; UE: Unerwünschtes Ereignis; UESI: Unerwünschtes Ereignis von besonderem Interesse; - : nicht berechenbar.

### Schwere unerwünschte Ereignisse (CTCAE-Grad $\geq 3$ )

Ein signifikanter Unterschied zeigte sich für „Diarrhö“ (PT) bzw. „Erkrankungen des Gastrointestinaltrakts“ (SOC) zuungunsten für Nirogacestat.

Tabelle 24: Schwere UE (CTCAE-Grad  $\geq 3$ ), die bei  $\geq 5$  % der Personen in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022)

Studie DeFi Schwere UE (CTCAE-Grad $\geq 3$ ) <sup>1)</sup> MedDRA-Systemorganklasse <sup>2)</sup> Preferred Term	Nirogacestat N = 69 <sup>3)</sup> Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)	Placebo N = 72 <sup>3)</sup> Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)	HR <sup>4)</sup> [95%-KI]; p-Wert <sup>5)</sup>
<b>Erkrankungen des Gastrointestinaltrakts</b>	n. a. [n. a.; n. a.] <b>16 (23)</b>	n. a. [n. a.; n. a.] <b>4 (6)</b>	<b>4,56<sup>6)</sup> [1,52; 13,7]; 0,002</b>
Diarrhö	n. a. [n. a.; n. a.] 11 (16)	n. a. [21,3; n. a.] 1 (1)	11,5 <sup>7)</sup> [1,48; 89,5]; 0,002
<b>Infektionen und parasitäre Erkrankungen</b>	n. a. [n. a.; n. a.] <b>10 (14)</b>	n. a. [n. a.; n. a.] <b>6 (8)</b>	<b>1,48 [0,54; 4,08] 0,22</b>
<b>Stoffwechsel- und Ernährungsstörungen<sup>8)</sup></b>	k. A. [k. A.; k. A.] <b>6 (9)</b>	n. a. [n. a.; n. a.] <b>0 (0)</b>	- <sup>9)</sup>
<b>Erkrankungen der Haut und des Unterhautgewebes<sup>8)</sup></b>	k. A. [k. A.; k. A.] <b>6 (9)</b>	n. a. [n. a.; n. a.] <b>0 (0)</b>	- <sup>9)</sup>
Makulopapulöser Ausschlag <sup>8)</sup>	k. A. [k. A.; k. A.] 4 (6)	n. a. [n. a.; n. a.] 0 (0)	- <sup>9)</sup>
<b>Untersuchungen<sup>8)</sup></b>	k. A. [k. A.; k. A.] <b>5 (7)</b>	k. A. [k. A.; k. A.] <b>6 (8)</b>	k. A.

<sup>1)</sup> Definiert als jedes unerwünschte medizinische Ereignis mit anfänglichem Auftreten oder zunehmenden Schweregrad nach Einnahme der ersten Dosis der Studienbehandlung bis 30 Tage nach der letzten Dosis der Studienmedikation. Wenn eine Person ein kodiertes Sicherheitsereignis mehr als einmal erlebte, wurde dieses Ereignis nur einmal gezählt.

<sup>2)</sup> Die Kodierung erfolgte nach MedDRA-Version 24.0.

<sup>3)</sup> Die Sicherheitspopulation umfasst alle Personen, die zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben.

<sup>4)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation. Die Referenzbehandlung war Placebo.

<sup>5)</sup> Basierend auf einseitigem stratifiziertem Log-Rank-Test mit Placebo als Referenz ( $\alpha = 0,025$ ).

<sup>6)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,03$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>7)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,049$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>8)</sup> Angaben gemäß Modul 5; im Dossier (Modul 4) liegen keine Angaben vor. Es liegen keine Ereigniszeitanalysen vor.

<sup>9)</sup> Die Berechnung eines HR konnte aufgrund fehlender Ereignisse im Placebo-Arm nicht erfolgen und Log-Rank-Test wurde nicht berichtet.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; DB: doppelblind; HR: Hazard Ratio; k. A.: keine Angabe; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; n. a.: nicht anwendbar; UE: Unerwünschtes Ereignis; - : nicht berechenbar.

### Schwerwiegende unerwünschte Ereignisse

Tabelle 25: SUE, die bei  $\geq 5\%$  der Personen in mindestens einem Studienarm aufgetreten sind, nach Systemorganklasse und Preferred Term, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), Sicherheitspopulation (Datenschnitt: 07.04.2022)

Studie DeFi SUE <sup>1)</sup> MedDRA-Systemorganklasse <sup>2)</sup>	Nirogacestat N = 69 <sup>3)</sup> Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)	Placebo N = 72 <sup>3)</sup> Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)	HR <sup>4)</sup> [95%-KI]; p-Wert <sup>5)</sup>
<b>Infektionen und parasitäre Erkrankungen</b>	n. a. [n. a.; n. a.] 4 (6)	n. a. [n. a.; n. a.] 5 (7)	0,61 [0,16; 2,29]; 0,23
<b>Erkrankungen der Geschlechts- organe und der Brustdrüse</b>	n. a. [n. a.; n. a.] 4 (6)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>6)</sup>

<sup>1)</sup> Definiert als jedes unerwünschte medizinische Ereignis mit anfänglichem Auftreten oder zunehmenden Schweregrad nach Einnahme der ersten Dosis der Studienbehandlung bis 30 Tage nach der letzten Dosis der Studienmedikation. Wenn eine Person ein kodiertes Sicherheitsereignis mehr als einmal erlebte, wurde dieses Ereignis nur einmal gezählt.

<sup>2)</sup> Die Kodierung erfolgte nach MedDRA-Version 24.0.

<sup>3)</sup> Die Sicherheitspopulation umfasst alle Personen, die zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben.

<sup>4)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>5)</sup> Basierend auf einseitigem stratifiziertem Log-Rank-Test mit Placebo als Referenz ( $\alpha = 0,025$ ).

<sup>6)</sup> Die Berechnung eines HR konnte aufgrund fehlender Ereignisse im Placebo-Arm nicht erfolgen und Log-Rank-Test wurde nicht berichtet.

Abkürzungen: DB: doppelblind; HR: Hazard Ratio; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; n. a.: nicht anwendbar; SUE: Schwerwiegendes unerwünschtes Ereignis; - : nicht berechenbar.

**Unerwünschte Ereignisse von besonderem Interesse**
**Tabelle 26: UESI; Studie DeFi (DB-Phase) Sicherheitspopulation (Datenschnitt: 07.04.2022)**

<b>Studie DeFi UESI<sup>1)2)</sup></b>	<b>Nirogacestat N = 69 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>Placebo N = 72 Mediane Zeit bis Ereignis, Monate [95%-KI] n (%)</b>	<b>HR [95%-KI]<sup>3)</sup>; p-Wert<sup>4)</sup></b>
<b>Störung des Fortpflanzungssystems<sup>5)</sup></b>			
<b>Amenorrhö<sup>6)</sup></b>			
Grad ≥ 2 <sup>7)</sup>	n. a. [n. a.; n. a.] 2 (3)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
Grad ≥ 3	n. a. [n. a.; n. a.] 0 (0)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
<b>Vorzeitige Menopause<sup>6)</sup></b>			
Grad ≥ 2 <sup>9)7)</sup>	n. a. [n. a.; n. a.] 11 (16)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
<b>Leberenzyme<sup>5)</sup></b>			
<b>Alanin-Aminotransferase erhöht<sup>6)</sup></b>			
Grad ≥ 2 <sup>7)</sup>	n. a. [n. a.; n. a.] 4 (6)	n. a. [21,4; n. a.] 2 (3)	1,73 <sup>10)</sup> [0,31; 9,76]; 0,27
Grad ≥ 3	n. a. [n. a.; n. a.] 2 (3)	n. a. [21,4; n. a.] 1 (1)	1,55 [0,13; 18,1]; 0,36
<b>Aspartat-Aminotransferase erhöht<sup>6)</sup></b>			
Grad ≥ 2 <sup>7)</sup>	n. a. [n. a.; n. a.] 3 (4)	n. a. [21,4; n. a.] 2 (3)	1,18 <sup>11)</sup> [0,19; 7,46]; 0,43
Grad ≥ 3 <sup>7)</sup>	n. a. [n. a.; n. a.] 2 (3)	n. a. [n. a.; n. a.] 1 (1)	2,15 [0,20; 23,7]; 0,26
<b>Alkalische Phosphatase erhöht</b>	k. A.	k. A.	k. A.
<b>Elektrolytinsuffizienz<sup>12)</sup></b>			
<b>Hypophosphatämie (Grad ≥ 3)<sup>6)7)</sup></b>	n. a. [n. a.; n. a.] 2 (3)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
<b>Hypokaliämie</b>	k. A.	k. A.	k. A.
<b>Hypomagnesiämie</b>	k. A.	k. A.	k. A.
<b>Hautausschlag<sup>13)</sup></b>			
<b>Folikulitis<sup>6)</sup></b>			
Grad ≥ 2, wenn klinisch signifikant <sup>7)</sup>	n. a. [n. a.; n. a.] 6 (9)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
Grad ≥ 3	n. a. [n. a.; n. a.] 4 (6)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
<b>Hidradenitis suppurativa<sup>6)</sup></b>			
Grad ≥ 2, wenn klinisch signifikant <sup>7)</sup>	n. a. [n. a.; n. a.] 5 (7)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>
Grad ≥ 3	n. a. [n. a.; n. a.] 1 (1)	n. a. [n. a.; n. a.] 0 (0)	_ <sup>8)</sup>

<b>Studie DeFi UESI<sup>1)2)</sup></b>	<b>Nirogacestat N = 69</b> <i>Mediane Zeit bis Ereignis, Monate [95%-KI]</i> n (%)	<b>Placebo N = 72</b> <i>Mediane Zeit bis Ereignis, Monate [95%-KI]</i> n (%)	<b>HR [95%-KI]<sup>3)</sup>; p-Wert<sup>4)</sup></b>
<b>Makulopapulöser Ausschlag<sup>6)</sup></b>			
Grad $\geq 2$ , wenn klinisch signifikant <sup>7)</sup>	k. A. [k. A.; k. A.] 8 (12)	k. A. [k. A.; k. A.] 0 (0)	k. A.
Grad $\geq 3$	n. a. [n. a.; n. a.] 4 (6)	n. a. [n. a.; n. a.] 0 (0)	... <sup>8)</sup>
<b>Pruritischer Ausschlag</b>	k. A.	k. A.	k. A.
<b>Erythematöser Ausschlag</b>	k. A.	k. A.	k. A.
<b>Arzneimittelreaktionen</b>			
<b>Allergische Reaktion</b>	k. A.	k. A.	k. A.
<b>Anaphylaxie</b>	k. A.	k. A.	k. A.

<sup>1)</sup> Ausgewählte UE, die unabhängig von ihrem Zusammenhang mit der Studienbehandlung gemeldet wurden.

<sup>2)</sup> Die Kodierung erfolgte nach MedDRA-Version 24.0.

<sup>3)</sup> Die Berechnung des HR erfolgte anhand eines stratifizierten Cox-Proportional-Hazard-Modells. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgte nach der Tumorlokalisation.

<sup>4)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ); p-Werte unter 0,025 werden als statistisch signifikant angesehen werden.

<sup>5)</sup> UESI umfasst alle UE-Grad  $\geq 2$  gemäß CTCAE-Version 5.0.

<sup>6)</sup> Kein SUE aufgetreten.

<sup>7)</sup> Entspricht allen UESI jeglichen Schweregrads.

<sup>8)</sup> Aufgrund fehlender Ereignisse im Placebo-Arm wurde kein HR berechnet.

<sup>9)</sup> Alle UE entsprachen dem CTCAE-Grad 2.

<sup>10)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,03$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>11)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,04$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

<sup>12)</sup> UESI umfasst alle UE Grad  $\geq 3$  gemäß CTCAE-Version 5.0.

<sup>13)</sup> UESI umfasst alle UE, wenn klinisch signifikant Grad 2 und alle Grade  $\geq 3$  gemäß CTCAE-Version 5.0.

Abkürzungen: CTCAE: Common Terminology Criteria for Adverse Events; DB: doppelblind; HR: Hazard Ratio; k. A.: keine Angabe; KI: Konfidenzintervall; MedDRA: Medical Dictionary for Regulatory Activities; n. a.: nicht anwendbar; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis; UESI: Unerwünschtes Ereignis von besonderem Interesse; - : nicht berechenbar.

## 4 Diskussion der Methodik und Ergebnisse

### 4.1 Design und Methodik der Studie

#### Studiendesign

Die Nutzenbewertung stützt sich auf die pivotale Studie DeFi, eine multizentrische, randomisierte, doppelblinde, placebokontrollierte Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit bei erwachsenen Personen mit fortschreitenden Desmoidtumoren, die eine systemische Therapie erfordern. Die eingeschlossenen Personen wurden stratifiziert nach der primären Tumorklassifikation (intra-abdominal vs. extra-abdominal) randomisiert, wobei Personen mit mehreren Zieltumoren, die sich sowohl intra- als auch extra-abdominal befanden, als intra-abdominal klassifiziert wurden. Auf eine bis zu 28-tägige Screening-Phase folgte die Randomisierung an Zyklus 1, Tag 1 mit der beginnenden DB-Phase (28-Tage-Zyklen) sowie eine Follow-up-Phase (30 Tage nach der letzten Dosierung) und eine sich anschließende offene OLE-Phase, in der alle Personen mit 150 mg Nirogacestat behandelt wurden (bis zu 5 Jahre). Die Personen konnten so lange in der DB-Phase bleiben, bis u. a. ein radiographischer Progress mittels zentraler Bildauswertung oder eine klinische Progression durch das Prüfpersonal festgestellt wurde. Im Fall, dass eine radiographische Progression (gemäß RECIST) mittels zentraler Bildauswertung festgestellt wird, erfolgt eine Entblindung und die Personen hatten die Möglichkeit die optionale OLE-Phase zu beginnen. Demgegenüber erfolgt für Personen, bei denen eine klinische Progression durch das Prüfpersonal, aber keine radiographische Progression festgestellt wurde, keine Entblindung und für die Personen bestand keine Möglichkeit in die OLE-Phase überzugehen.

Es erfolgten 5 globale Amendments des Originalprotokolls nach Einschluss der ersten Studiensperson. Relevante Protokolländerungen umfassten das Hinzufügen und Entfernen von UESI (siehe Tabelle 3). Zudem wurde der primäre Endpunkt „PFS“ um die klinische Progression erweitert, definiert als jegliche qualifizierte klinische Progression oder Tod.

Die Studie ist abgeschlossen (finale Datenbanksperrung der OLE-Phase: 19.12.2024). Für die Nutzenbewertung von Nirogacestat liegt der finale Datenschnitt der DB-Phase vom 07.04.2022 vor. Der Datenschnitt vom 07.04.2022 wird als geeignet für die Nutzenbewertung angesehen.

#### Studienpopulation

In der Studie DeFi wurden zwischen 2019 und 2021 insgesamt 142 Personen an 52 Studienzentren in 7 Ländern rekrutiert. Es wurden erwachsene Personen  $\geq 18$  Jahre mit histologisch bestätigten Desmoidtumoren eingeschlossen, die innerhalb von 12 Monaten vor der Screening-Visite eine Progression von  $\geq 20\%$  gemäß RECIST vorweisen. Zudem mussten die Personen entweder behandlungsnaive oder rezidivierende oder refraktäre, messbar fortschreitende Desmoidtumore aufweisen, die nicht, ohne erhebliche Morbidität, operativ behandelt werden können.

Die Studienpopulation bestand überwiegend aus jungen Erwachsenen mit einem medianen Alter von etwa 34 Jahren in beiden Studienarmen. Die Geschlechtsverteilung zeigte eine erwartbare Prädominanz weiblicher Personen (64 % unter Nirogacestat vs. 65 % unter Placebo). Es wurden geringfügige Ungleichgewichte in den Merkmalen „Abstammung“ und „Region“ beobachtet. Bezogen auf die krankheitsspezifischen Merkmale zeigt sich, dass die Mehrheit der Personen in beiden Studienarmen, jedoch anteilig im Placebo-Arm höher, refraktär gegenüber einer vorherigen Desmoidtumor-Behandlung waren (61 % vs. 76 %). Gemäß gewichtetem DTSS-Gesamtscore war die mediane symptomatische Belastung zu Baseline im Nirogacestat- und Placebo-Arm vergleichbar (3,2 Punkte auf einer 11-Punkte-NRS), was auf eine geringere symptomatische Belastung hindeutet. Die Anzahl der Zieltumore war in beiden Studienarmen vergleichbar (Median:  $n = 1$ ). Unterschiede zeigten sich in der Größe der Zieltumore zu Baseline gemäß RECIST, wonach die mediane Größe im Placebo- größer als im Nirogacestat-Arm war (ca.

91 mm vs. ca. 116 mm). Die Mehrzahl der Personen im Placebo-Arm hatte eine Operation als Vortherapie (61 %), im Vergleich dazu waren es im Nirogacestat-Arm 44 %. Die vergleichbare Verteilung der Baseline- und krankheitsspezifischen Charakteristika bzw. die nur vereinzelt vermutlich zufallsbedingt aufgetretenen Unterschiede zwischen beiden Studienarmen weisen auf eine adäquate Randomisierung hin.

### **Studienmedikation**

Die Personen nahmen Fachinformation-konform zweimal täglich (ungefähr alle 12 Stunden, unabhängig von der Nahrungsaufnahme) 150 mg (aufgeteilt in 3 × 50 mg) Nirogacestat- bzw. Placebo-Tabletten ein (Tabelle 4). Die Behandlung erfolgte kontinuierlich in 28-Tage-Zyklen.

Die mediane Behandlungsdauer lag im Nirogacestat-Arm mit 20,6 Monaten (min; max: 0,3; 33,6) über der medianen Behandlungsdauer von 11,4 Monaten (min; max: 0,2; 32,5) im Placebo-Arm. Einen vorzeitigen Behandlungsabbruch während der DB-Phase wiesen mehr Personen im Nirogacestat- als im Placebo-Arm auf (30 % vs. 19 %). Der Grund hierfür war primär das Auftreten von UE. Ein gleiches Bild zeigte sich für die Personen, die die Studie vorzeitig abbrachen. Auch hier war der Anteil an Personen im Nirogacestat- höher als im Placebo-Arm (43 % vs. 29 %). Gründe im Nirogacestat-Arm waren hauptsächlich das Auftreten von UE (16 %) und das Auftreten einer radiographischen Progression (10 %), während der Hauptgrund im Placebo-Arm die Entscheidung der Patientinnen/Patienten (10 %) und die qualifizierte klinische Progression war (8 %).

Dosisanpassungen (Reduzierungen/Unterbrechungen) waren unter Nirogacestat häufiger als im Placebo-Arm: Dosisreduktionen erfolgten bei 42 % der Personen (vs. 0 % unter Placebo) und Dosisunterbrechungen erfolgten bei 65 % der Personen (vs. 17 % unter Placebo).

## **4.2 Zulassungsstatus und Zulassungspopulation**

### **Zulassungsstatus und Anwendungsgebiet**

Nirogacestat (Ogsiveo®) ist zugelassen für die Behandlung von erwachsenen Patienten mit fortschreitenden Desmoidtumoren als Monotherapie, die eine systemische Behandlung erfordern. Die zugelassene Dosierung von 150 mg wird zweimal täglich morgens und abends oral eingenommen. Die Behandlung soll bis zum Fortschreiten der Erkrankung oder bis zu einer inakzeptablen Toxizität fortgeführt werden [17]. Die Behandlung erfolgte gemäß den Angaben in der Fachinformation.

### **Übertragbarkeit auf deutschen Versorgungskontext**

Die Studie wurde in 7 Ländern an 52 Zentren durchgeführt. Von 142 Personen der Gesamtstudienpopulation nahmen 45 (32 %) an europäischen Standorten teil, darunter 10 Personen (7 %) in Deutschland. Die Mehrheit der Studienpopulation (68 %) verteilt sich auf nordamerikanische Standorte. Von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist somit auszugehen.

### 4.3 Zusammenfassende Beurteilung zum Verzerrungspotential und zur Ergebnissicherheit

Zusammenfassend zeigt die Studie DeFi, aufgrund ihres randomisierten und doppelblinden Designs, ein geringes Verzerrungspotential auf Studienebene. Die Ergebnissicherheit wird jedoch eingeschränkt durch die kurze Nachbeobachtungsdauer von nur 30 Tagen nach Studien- bzw. Behandlungsabbruch, was die langfristige Erfassung von Morbidität und Lebensqualität erschwert.

Auf Endpunktebene, insbesondere für PRO und Sicherheit, wird das Verzerrungspotential jedoch als hoch eingeschätzt. Gründe hierfür sind:

- Abnehmende Rücklaufquoten, die bereits ab Zyklus 4 unter 70 % liegen und zu einem hohen Anteil fehlender Werte in späteren Zyklen führen,
- fehlende Angaben zu Zensierungsgründen,
- fehlende Nachbeobachtung nach Behandlungsabbruch, die zu potentiell informativen Zensierungen führt.

Inwieweit es sich bei den Studienergebnissen um nachhaltige Effekte der dauerhaften Einnahme von Nirogacestat handelt, kann nicht abschließend beurteilt werden. Basierend auf den vorliegenden Daten können nur kurz- und mittelfristige Effekte von Nirogacestat bewertet werden. Daten aus einer OLE-Phase wären aufgrund fehlender Kontrollen nicht hilfreich für die weitere Beurteilung der gezeigten Effekte. Von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist auszugehen.

### 4.4 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserfassung dokumentiert. Während der doppelblinden Studienphase, bzw. bis zum Datenschnitt vom 07.04.2022, wurde ein Todesfall im Placebo-Arm berichtet.

### 4.5 Morbidität

Für die Endpunktkategorie „Morbidität“ legte der pU in Modul 4 Daten zu den Endpunkten „PFS“, „ORR“, „Veränderung des Tumolvolumens“ und zu patientenberichteten Endpunkten anhand der Instrumente „GODDESS-DTSS/-DTIS“, „BPI-SF“, „EORTC QLQ-C30“, „PROMIS PF Short Form 10a/13“, „PGIS“ und „PGIC“ vor. Die detaillierte Bewertung der Endpunkte findet sich in Kapitel 2.3. Initial hat der pU die Endpunkte „PROMIS PF Short Form 10a/13“ und „GODDESS-DTIS“ der Endpunktkategorie „Lebensqualität“ zugeordnet. Da in diesen Endpunkten nur Teilaspekte des multidimensionalen Konstrukts der gesundheitsbezogenen Lebensqualität abgebildet wurden, erfolgte die Zuordnung zur Endpunktkategorie „Morbidität“.

Zusammenfassend wurden für die Nutzenbewertung die Ergebnisse für die patientenberichteten Endpunkte anhand der Instrumente „GODDESS DTSS/-DTIS“, „BPI-SF“, „EORTC QLQ-C30“ und „PGIC“ herangezogen.

Für die GODDESS-DTSS werden in der vorliegende Nutzenbewertung der „Gewichtete DTSS-Gesamtscore“ und der Domänenscore „Intra-abdominale Symptome“ berücksichtigt. Bei den Responderanalysen für „Zeit bis zur ersten Verschlechterung“ zeigte sich ein signifikanter Unterschied für den gewichteten DTSS-Gesamtscore (HR: 0,50 (95%-KI: [0,30; 0,84])) zwischen den beiden Behandlungsarmen zugunsten von Nirogacestat. Für den Domänenscore „Intra-abdominale Symptome“ zeigte sich kein statistisch signifikanter Unterschied.

Die GODDESS-DTIS wurden mit den 3 Domänenscores „Körperliche Funktion“, „Schlaf“ und „Emotion“ im Rahmen der vorliegenden Nutzenbewertung berücksichtigt. In den Domänenscores „Körperliche Funktion“ (HR: 0,33 (95%-KI: [0,18; 0,60])) und „Schlaf“ (HR: 0,57 (95%-KI: [0,34; 0,94])) zeigte sich für die Zeit bis zur ersten Verschlechterung ein signifikanter Unterschied zwischen den Behandlungsarmen zugunsten von Nirogacestat; für den Domänenscore „Emotion“ zeigte sich kein statistisch signifikanter Unterschied. In der Subgruppenanalyse zeigte sich eine statistisch signifikante Interaktion für den Domänenscore „Körperliche Funktion“ für die Variablen „Tumorklassifikation“ und „Jegliche vorherige Therapie“: Bei Personen mit extra-abdominaler Manifestation ergab sich ein signifikanter Unterschied zugunsten von Nirogacestat (HR: 0,13 (95%-KI: [0,05; 0,32])), während bei intra-abdominaler Manifestation kein signifikanter Unterschied (HR: 1,58 (95%-KI: [0,56; 4,48])) gezeigt werden konnte; bei Personen mit jeglicher vorherigen Therapie ergab sich ein signifikanter Unterschied zugunsten von Nirogacestat (HR: 0,17 (95%-KI: [0,07; 0,41])), während bei keiner vorherigen Therapie kein signifikanter Unterschied (HR: 0,72 (95%-KI: [0,23; 2,25])) gezeigt werden konnte. Statistisch signifikante Interaktionen liegen jedoch nicht für andere Morbiditätseindpunkte vor, sodass die Relevanz dieser Subgruppeneffekte insgesamt unklar ist.

Für den BPI-SF wurden 2 Skalen für die Nutzenbewertung herangezogen. Im Ergebnis zeigte sich für die Skala „Beeinträchtigung durch den Schmerz“ ein statistisch signifikanter Unterschied zwischen den beiden Behandlungsarmen zugunsten von Nirogacestat (HR: 0,42 [95%-KI: 0,23; 0,96]). Für die Skala „Schmerzintensität“ zeigte sich kein statistisch signifikanter Unterschied.

Die Responderanalyse „Zeit bis zu jeglicher Verschlechterung“ des allgemeinen Gesundheitszustands gemessen mittels PGIC zeigt einen signifikanten Unterschied zugunsten von Nirogacestat (HR: 0,53 (95%-KI: [0,33; 0,86])). Die Verletzung der Proportional-Hazard-Annahme war statistisch signifikant ( $p = 0,002$ ).

Die Responderanalysen „Zeit bis zur ersten Verschlechterung“ der Symptomskalen des EORTC QLQ-C30 zeigen für die Skalen „Übelkeit und Erbrechen“ (HR: 3,00 (95%-KI: [1,80; 5,00])), „Appetitverlust“ (HR: 1,96 (95%-KI: [1,13; 3,41])) sowie „Diarrhö“ (HR: 3,58 (95%-KI: [2,20; 5,84])) einen statistisch signifikanten Unterschied zuungunsten von Nirogacestat. In der Symptomskala „Obstipation“ (HR: 0,49 (95%-KI: [0,26; 0,92])) zeigte sich hingegen ein statistisch signifikanter Unterschied zugunsten von Nirogacestat. Die Verletzung der Proportional-Hazard-Annahme war für die Symptomskala „Übelkeit und Erbrechen“ statistisch signifikant ( $p = 0,03$ ).

In der Gesamtschau zeigen sich Vorteile hinsichtlich der erstmaligen Verschlechterung von Desmoidtumor-bedingten Symptomen (GODDESS-DTSS), der körperlichen Funktion (GODDESS-DTIS), des Schlafs (GODDESS-DTIS), Obstipationen (EORTC QLQ-C30), Beeinträchtigung durch den Schmerz (BPI-SF). Diese Einschätzung wird durch die Ergebnisse zum allgemeinen Gesundheitszustand, gemessen mittels PGIC, bestätigt. Jedoch ist hinsichtlich der Effektschätzer für den allgemeinen Gesundheitszustand, gemessen mittels PGIC, und der Symptomskala „Übelkeit und Erbrechen“ des EORTC QLQ-C30 die statistisch signifikante Verletzung der Proportional-Hazard-Annahme anzumerken. Sie legt nahe, dass die Größe des Effektschätzers über die Zeit variiert, was sich auch in den Kaplan-Meier-Kurven widerspiegelt. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Der Vorteil hinsichtlich der körperlichen Funktion zeigte sich über verschiedene Erhebungsinstrumente hinweg, sowohl im krankheitsübergreifenden EORTC QLQ-C30 (siehe Lebensqualität) als auch im krankheitsspezifischen Instrument GODDESS-DTIS (siehe Morbidität). Gleichzeitig zeigten sich Effekte zuungunsten von Nirogacestat, hinsichtlich „Übelkeit und Erbrechen“, „Appetitverlust“ sowie „Diarrhö“. Dies könnte Nebenwirkungen von Nirogacestat widerspiegeln, die sich auch in den Analysen der Sicherheitsendpunkte zeigen.

Das Verzerrungspotential für die beschriebenen Morbiditätspunkte wird insgesamt als hoch bewertet, da zum einen die Rücklaufquoten bereits zu bzw. ab Zyklus 4 bei unter 70 % lagen und zum anderen bei frühzeitigem Behandlungs-/Studienabbruch Personen i. d. R. lediglich 30 Tage nachbeobachtet wurden. Entsprechend ist auf eine potentielle Verzerrung durch informative Zensierung hinzuweisen. Zudem ist anzumerken, dass einige Symptome (z. B. Schmerz, Fatigue, Schlaf) über mehrere Erhebungsinstrumente (GODDESS-DTSS/-DTIS, BPI-SF, EORTC QLQ-C30) vermeintlich doppelt erfasst und somit ausgewertet wurden.

#### 4.6 Lebensqualität

In den Responderanalysen (Zeit bis zur ersten Verschlechterung) der Funktionsskalen und der globalen Skala „Allgemeiner Gesundheitsstatus / Lebensqualität“ des EORTC QLQ-C30 zeigte sich ein signifikanter Unterschied zwischen den Behandlungsarmen in der Subskala „Körperliche Funktion“ zugunsten von Nirogacestat (HR: 0,38 (95%-KI: [0,21; 0,70])). Auch in der Endpunktkategorie „Morbidität“ zeigte sich für die krankheitsspezifische Subdomäne „Körperliche Funktion“ der GODDESS-DTIS ein gleichgerichteter Effekt. Bei beiden Skalen stehen Einschränkungen bei der Durchführung alltäglicher Aktivitäten im Fokus. Insgesamt zeigen sich in der Endpunktkategorie „Lebensqualität“ lediglich Unterschiede in einem Teilbereich der gesundheitsbezogenen Lebensqualität, der wesentlich durch die Symptomatik bestimmt wird. Unterschiede in der sozialen oder psychischen Dimension sind aus den vorgelegten Daten nicht abzuleiten.

Das Verzerrungspotential für die Lebensqualität wird insgesamt als hoch bewertet, da zum einen die Rücklaufquoten bereits zu bzw. ab Zyklus 4 bei unter 70 % lagen und zum anderen bei frühzeitigem Behandlungs-/Studienabbruch Personen i. d. R. lediglich 30 Tage nachbeobachtet wurden. Entsprechend ist auf eine potentielle Verzerrung durch informative Zensierung hinzuweisen.

#### 4.7 Sicherheit

Die Erfassung aller UE erfolgte ab dem Zeitpunkt der Unterzeichnung der Einverständniserklärung bis 30 Tage nach Gabe der letzten Studiendosis. Zusätzlich erfolgte bei Frauen mit UE im Kontext einer primären Ovarialinsuffizienz und/oder Amenorrhö die Messung des Hormonspiegels alle 3 Monate bis zum Ende des Ereignisses bzw. für mindestens 90 Tage nach Abbruch der Studienbehandlung.

Die mediane Behandlungsdauer im Nirogacestat-Arm betrug 20,6 Monate (min; max: 0,3; 33,6) und lag somit über der medianen Behandlungsdauer im Placebo-Arm (11,4 Monate (min; max: 0,2; 32,5)).

Für die Gesamtrate von UE CTCAE-Grad  $\geq 3$  (HR: 3,80 (95%-KI: [1,98; 7,31])) und UE, die zum Abbruch der Studienmedikation führten (HR: 8,26 (95%-KI: [1,90; 36,0])), wurden statistisch signifikante Unterschiede zuungunsten von Nirogacestat registriert. Ergänzend ist für die UE, die zum Abbruch der Studienmedikation führten, die statistisch signifikante Verletzung der Proportional-Hazard-Annahme anzumerken. Sie legt nahe, dass die Größe des Effektschätzers über die Zeit variiert, was sich auch in den Kaplan-Meier-Kurven widerspiegelt. Die statistische Signifikanz des Effekts, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Für UE mit Inzidenz  $\geq 10$  % zeigten sich in den Ereigniszeitanalysen statistisch signifikante Unterschiede zuungunsten von Nirogacestat für die SOC „Erkrankungen des Gastrointestinaltrakts“, „Infektionen und parasitäre Erkrankungen“, Stoffwechsel- und Ernährungsstörungen“, „Erkrankungen der Geschlechtsorgane und der Brustdrüse“, „Erkrankungen der Atemwege, des Brustraums und Mediastinums“, „Erkrankungen der Haut und des Unterhautgewebes“ sowie „Gefäßerkrankungen“. Für die SOC „Gutartige, bösartige und unspezifische Neubildungen (einschl. Zysten und Polypen)“ zeigten sich hingegen statistisch signifikante Unterschiede zugunsten von Nirogacestat. Bezogen

auf schwere UE (CTCAE-Grad  $\geq 3$ ) zeigte sich ein signifikanter Unterschied für „Diarrhö“ (PT) zuungunsten für Nirogacestat. Für SUE der SOC „Infektionen und parasitäre Erkrankungen“ zeigte sich zwischen den Behandlungsarmen kein signifikanter Unterschied. In der SOC „Erkrankungen der Geschlechtsorgane und der Brustdrüse“ konnte aufgrund fehlender Ereignisse in einem Arm (Placebo) kein HR berechnet werden. Im Ergebnis zeigte sich ein Unterschied zwischen den Gruppen mit 4 Personen (6 %) im Nirogacestat-Arm im Vergleich zu keiner Person im Placebo-Arm. Für die UESI, bei denen ein HR berechnet werden konnte, zeigten sich zwischen den Behandlungsarmen keine signifikanten Unterschiede.

Zusammenfassend zeigten sich im Nirogacestat-Arm hinsichtlich der Gesamtraten von UE CTCAE-Grad  $\geq 3$  und UE, die zum Abbruch der Studienmedikation führten, statistisch signifikante Unterschiede zuungunsten von Nirogacestat. Bezogen auf schwere UE (CTCAE-Grad  $\geq 3$ ) zeigte sich ein signifikanter Unterschied für „Diarrhö“ (PT) zuungunsten für Nirogacestat. Zu beachten ist hierbei allerdings der relative Vergleich gegenüber Placebo. Darüber hinaus bleibt zu beachten, dass bei frühzeitigem Behandlungs-/Studienabbruch Personen i. d. R. lediglich 30 Tage nachbeobachtet wurden. Entsprechend wird das Verzerrungspotential aufgrund einer potentiell informativen Zensierung als hoch eingeschätzt.

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

Die Angaben der Fachinformation sind zu berücksichtigen.

## 6 Zusammenfassung der Nutzenbewertung

Nirogacestat ist zugelassen für Behandlung von erwachsenen Patienten mit fortschreitenden Desmoidtumoren als Monotherapie, die eine systemische Behandlung erfordern. Die Nutzenbewertung von Nirogacestat basiert auf der zulassungsbegründenden Studie DeFi, eine multi-zentrische, randomisierte, doppelblinde, placebokontrollierte, ereignisgesteuerte Phase-III-Studie.

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 Nirogacestat 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 27: Zusammenfassende Darstellung der Ergebnisse der Studie DeFi (DB-Phase)

Studie DeFi Darstellung der Ergebnisse	Nirogacestat N = 70		Placebo N = 72		Nirogacestat vs. Placebo	Effekt
<b>Mortalität</b>						
<b>Todesfälle<sup>1)</sup></b>	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	HR [95%-KI] <sup>2)</sup> ; p-Wert <sup>3)</sup>	
Gesamt- überleben	70	n. a. [n. a.; n. a.]; 0 (0)	72	n. a. [n. a.; n. a.]; 1 (1)	n. a. <sup>4)</sup>	
<b>Morbidität<sup>5)</sup></b>						
	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	HR [95%-KI]; p-Wert <sup>3)</sup>	
<b>GODDESS-DTSS – Zeit bis zur ersten Verschlechterung um <math>\geq 15</math> % der Skalenspannweite</b>						
Gewichtete DTSS- Gesamtscore <sup>6)</sup>	70	n. a. [10,9; n. a.]; 26 (37)	72	4,4 [2,37; n. a.]; 40 (56)	0,50 [0,30; 0,84]; 0,004	$\uparrow$
Intra-abdominale Symptome <sup>7)</sup>	70	n. a. [1,64; n. a.]; 12 (17)	72	3,52 [0,76; n. a.]; 15 (21)	0,95 [0,41; 2,22] 0,45	$\leftrightarrow$
<b>GODDESS DTIS – Zeit bis zur ersten Verschlechterung um <math>\geq 15</math> % Skalenspannweite</b>						
Körperliche Funktion	70	n. a. [n. a.; n. a.]; 16 (23)	72	8,54 [6,18; 14,7]; 35 (49)	0,33 [0,18; 0,6]; < 0,0001	$\uparrow$
Schlaf	70	n. a. [8,28; n. a.]; 28 (40)	72	7,36 [3,84; 11,1]; 39 (54)	0,57 [0,34; 0,94] 0,01	$\uparrow$
Emotion	70	n. a. [6,44; n. a.]; 30 (43)	72	4,67 [2,83; 17,5]; 37 (51)	0,61 [0,38; 1,01] 0,03	$\leftrightarrow$
<b>BPI-SF – Zeit bis zur ersten Verschlechterung um <math>\geq 15</math> % Skalenspannweite</b>						
Beeinträchtigung durch d. Schmerz	70	n. a. [n. a.; n. a.]; 12 (17)	72	n. a. [n. a.; n. a.]; 22 (31)	0,47 [0,23; 0,96] 0,02	$\uparrow$
Schmerz- intensität	70	n. a. [n. a.; n. a.]; 11 (16)	72	n. a. [n. a.; n. a.]; 19 (26)	0,50 [0,24; 1,06] 0,03	$\leftrightarrow$

Studie DeFi Darstellung der Ergebnisse	Nirogacestat N = 70		Placebo N = 72		Nirogacestat vs. Placebo	Effekt
<b>PGIC – Zeit bis zur ersten Verschlechterung um <math>\geq 15</math> % Skalenspannweite</b>						
	70	22,3 [3,75; n. a.]; 32 (46)	72	4,86 [3,48; 8,28]; 47 (65)	0,53 <sup>8)</sup> [0,33; 0,86] 0,004	↑
<b>EORTC QLQ-C30: Symptomskalen – Zeit bis zur ersten Verschlechterung um <math>\geq 10</math> Punkte</b>						
Obstipation	70	n. a. [n. a.; n. a.]; 16 (23)	72	n. a. [4,40; n. a.]; 27 (38)	0,49 [0,26; 0,92]; 0,01	↑
Übelkeit und Erbrechen	70	0,92 [0,92; 2,73]; 51 (73)	72	21,0 [9,20; n. a.]; 28 (39)	3,00 <sup>9)</sup> [1,80; 5,00]; < 0,0001	↓
Appetitverlust	70	6,44 [1,84; n. a.]; 37 (53)	72	n. a. [8,54; n. a.]; 23 (32)	1,96 [1,13; 3,41]; 0,01	↓
Diarrhö	70	0,92 [0,92; 1,84]; 55 (79)	72	9,23 [5,72; n. a.]; 32 (44)	3,58 [2,20; 5,84]; < 0,0001	↓
<b>EORTC QLQ-C30: Alle anderen Symptomskalen<sup>10)</sup></b>						
In keiner weiteren Symptomskala zeigte sich ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen (siehe Tabelle 20).						↔
<b>Lebensqualität<sup>5)</sup></b>						
	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	HR [95%-KI]; p-Wert <sup>3)</sup>	
<b>EORTC QLQ-C30: Funktionsskalen und globaler Gesundheitsstatus – Zeit bis zur ersten Verschlechterung um <math>\geq 10</math> Punkte</b>						
Körperliche Funktion	70	n. a. [n. a.; n. a.]; 16 (23)	72	8,38 [7,39; 18,8]; 33 (46)	0,38 [0,21; 0,70]; 0,001	↑
<b>EORTC QLQ-C30: Weitere Funktionsskalen und globaler Gesundheitsstatus<sup>11)</sup></b>						
In keiner weiteren Funktionsskala bzw. der globalen Skala „Gesundheitsstatus“ zeigte sich ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen (siehe Tabelle 21).						↔
<b>Sicherheit<sup>12)</sup></b>						
<b>Unerwünschte Ereignisse</b>	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	N	Mediane Zeit bis Ereignis, Monate [95%-KI]; Personen m. Ereignis, n (%)	HR [95%-KI]; p-Wert <sup>3)</sup>	
Schwere UE (CTCAE-Grad $\geq 3$ )	69	17,5 [3,32; 27,6] 38 (55)	72	n. a. [21,3; n. a.]; 12 (17)	3,80 [1,98; 7,31]; < 0,0001	↓
SUE	69	n. a. [n. a.; n. a.]; 14 (20)	72	n. a. [21,3; n. a.]; 8 (11)	1,77 <sup>13)</sup> [0,74; 4,26]; 0,10	↔
UE, das zum Abbruch der Studienmedi- kation führte	69	n. a. [n. a.; n. a.]; 16 (23)	72	n. a. [22,7; n. a.]; 2 (3)	8,26 <sup>13)</sup> [1,90; 36,0]; 0,0004	↓

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

<sup>2)</sup> Das HR wurde aus dem stratifizierten Cox-Proportional-Hazards-Modell geschätzt. Es wurde die „Exact Method for Ties“ verwendet. Die Stratifizierung erfolgt nach der Tumorlokalisation.

- 
- <sup>3)</sup> Der p-Wert stammt aus einem einseitigen stratifizierten Log-Rank-Test ( $\alpha = 0,025$ ). P-Werte unter 0,025 werden als statistisch signifikant angesehen.
- <sup>4)</sup> Die Berechnung eines adäquaten Effektschätzers (Hazard Ratio) war aufgrund fehlender Ereignisse im Interventionsarm nicht möglich.
- <sup>5)</sup> Für alle PRO: Eine Zensurierung erfolgte bei Personen, bei denen zum Zeitpunkt der Analyse keine Verschlechterung beobachtet wurde. Diese wurden zum Zeitpunkt der letzten verfügbaren PRO-Bewertung zum oder vor dem Datenschnitt der Analysen zensiert.
- <sup>6)</sup> Der „Gewichtete DTSS-Gesamtscore“ berücksichtigt die Items 1 bis 7, wobei die Schmerzskala (Item 1–3) gemittelt in die Berechnung des DTSS-Gesamtscore einbezogen wird. Die Items werden auf einer 11-stufigen numerischen Bewertungsskala von 0 („nicht vorhanden“) bis 10 („so schlimm, wie nur vorstellbar“) bewertet.
- <sup>7)</sup> Der Domänenscore „Intra-abdominale Symptome“ basiert auf den Items 9–11 der GODDESS-DTSS; die Beantwortung erfolgte nur durch Personen, welche die entsprechende Tumorlokalisation in Item 8 gewählt haben. Die Items werden auf einer 11-stufigen numerischen Bewertungsskala von 0 („nicht vorhanden“) bis 10 („so schlimm, wie nur vorstellbar“) bewertet.
- <sup>8)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,002$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effektes, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.
- <sup>9)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,03$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effektes, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.
- <sup>10)</sup> Fatigue, Schmerz, Dyspnoe und Schlaflosigkeit.
- <sup>11)</sup> Globaler Gesundheitsstatus, Kognitive Funktion, Emotionale Funktion, Rollenfunktion und Soziale Funktion.
- <sup>12)</sup> Die Sicherheitspopulation umfasst alle Personen, welche zu einem Behandlungsarm randomisiert wurden und mindestens eine Medikationsdosis erhalten haben.
- <sup>13)</sup> Es liegt eine statistisch signifikante Verletzung der Proportional-Hazards-Annahme ( $p = 0,02$ ) vor. Dies weist darauf hin, dass die Größe des Effektschätzers über die Zeit variiert. Die statistische Signifikanz des Effektes, getestet mittels Log-Rank-Test, wird dadurch nicht in Frage gestellt.

Abkürzungen: BPI-SF: Brief Pain Inventory – Short Form; CTCAE: Common Terminology Criteria for Adverse Events; DB: doppelblind; DTIS: Desmoid Tumor Impact Scale; DTSS: Desmoid Tumor Symptom Scale; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core Questionnaire; GODDESS: GOunder/DTRF DEsmoid Symptom/Impact Scale; HR: Hazard Ratio; KI: Konfidenzintervall; n. a.: nicht anwendbar; PGIC: Patient Global Impression of Change; PRO: Patientenberichteter Endpunkt; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

## Referenzen

1. **Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al.** New response evaluation criteria in solid tumours: revised RECIST guideline (vers. 1.1). *Eur J Cancer* 2009;45(2):228-47.
2. **European Medicines Agency (EMA).** Ogsiveo (nirogacestat): European public assessment report EMEA/H/C/006071/0000 [online]. 19.06.2025. Amsterdam (NED): EMA. [Zugriff: 15.10.2025]. URL: [https://www.ema.europa.eu/en/documents/assessment-report/ogsiveo-epar-public-assessment-report\\_en.pdf](https://www.ema.europa.eu/en/documents/assessment-report/ogsiveo-epar-public-assessment-report_en.pdf).
3. **Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, et al.** The EORTC QLQ-C30 Scoring Manual (3rd Edition) [online]. Brussels (BEL): European Organisation for Research and Treatment of Cancer; 2001. [Zugriff: 15.10.2025]. URL: <https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf>.
4. **Gounder M, Ratan R, Alcindor T, Schöffski P, van der Graaf WT, Wilky BA, et al.** Nirogacestat, a  $\gamma$ -secretase inhibitor for desmoid tumors. *N Engl J Med* 2023;388(10):898–912.
5. **Gounder MM, Atkinson TM, Bell T, Daskalopoulou C, Griffiths P, Martindale M, et al.** Gounder/Desmoid Tumor Research Foundation DEsmoid Symptom/Impact Scale (GODDESS((c))): psychometric properties and clinically meaningful thresholds as assessed in the phase 3 DeFi randomized controlled clinical trial. *Qual Life Res* 2023;32(10):2861-2873.
6. **Gounder MM, Maddux L, Paty J, Atkinson TM.** Prospective development of a patient-reported outcomes instrument for desmoid tumors or aggressive fibromatosis. *Cancer* 2020;126(3):531–539.
7. **Groenvold M, Klee MC, Sprangers MA, Aaronson NK.** Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J Clin Epidemiol* 1997;50(4):441–50.
8. **Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.** Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-766 e1.
9. **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.
10. **Norman GR, Stratford P, Regehr G.** Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50(8):869-79.
11. **Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE, Jr.** The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol* 2014;67(5):516–26.
12. **Sánchez Trullàs K, SpringWorks Therapeutics.** A randomized, double-blind, placebo-controlled, phase 3 trial of Nirogacestat versus placebo in adult patients with progressing desmoid tumors/aggressive fibromatosis (DT/AF); psychometric analysis report, NIR-DT-301 [unveröffentlicht]. 26.01.2022.

13. **SpringWorks Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Anhang 4-G; Nirogacestat (Ogsiveo), Behandlung erwachsener Patienten mit fortschreitenden Desmoidtumoren, die eine systemische Behandlung erfordern; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 15.10.2025.
14. **SpringWorks Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Anhang 4-H; Nirogacestat (Ogsiveo), Behandlung erwachsener Patienten mit fortschreitenden Desmoidtumoren, die eine systemische Behandlung erfordern; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 15.10.2025.
15. **SpringWorks Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Anhang 4-I; Nirogacestat (Ogsiveo), Behandlung erwachsener Patienten mit fortschreitenden Desmoidtumoren, die eine systemische Behandlung erfordern; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 15.10.2025.
16. **SpringWorks Therapeutics.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Nirogacestat (Ogsiveo), Behandlung erwachsener Patienten mit fortschreitenden Desmoidtumoren, die eine systemische Behandlung erfordern; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 15.10.2025.
17. **SpringWorks Therapeutics.** Ogsiveo 50 mg /100 mg / 150 mg Filmtabletten [online]. 08.2025. Frankfurt/Main. [Zugriff: 15.10.2025]. URL: <https://www.fachinfo.de>.
18. **SpringWorks Therapeutics.** A randomized, double-blind, placebo-controlled, phase 3 study of nirogacestat versus placebo in adult patients with progressing desmoid tumors/aggressive fibromatosis (DT); NIR-DT-301; clinical study report [unveröffentlicht]. 05.10.2023.
19. **SpringWorks Therapeutics.** A randomized, double-blind, placebo-controlled, phase 3 trial of nirogacestat versus placebo in adult patients with progressing desmoid tumors/aggressive fibromatosis (DT/AF); NIR-DT-30; protocol, last updated 09.02.2021 [unveröffentlicht]. 03.08.2018.
20. **SpringWorks Therapeutics.** A randomized, double-blind, placebo-controlled, phase 3 trial of nirogacestat versus placebo in adult patients with progressing desmoid tumors/aggressive fibromatosis (DT/AF); NIR-DT-301; statistical analysis plan [unveröffentlicht]. 07.04.2022.
21. **SpringWorks Therapeutics.** Statistical analysis plan for patient-reported outcomes: an addendum to NIR-DT-301 statistical analysis plan (PRO addendum); NIR-DT-301; protocol, v2 [unveröffentlicht]. 05.04.2022.

## Anhang

### EORTC QLQ-C30: Symptomskalen – Kaplan-Meier-Kurven

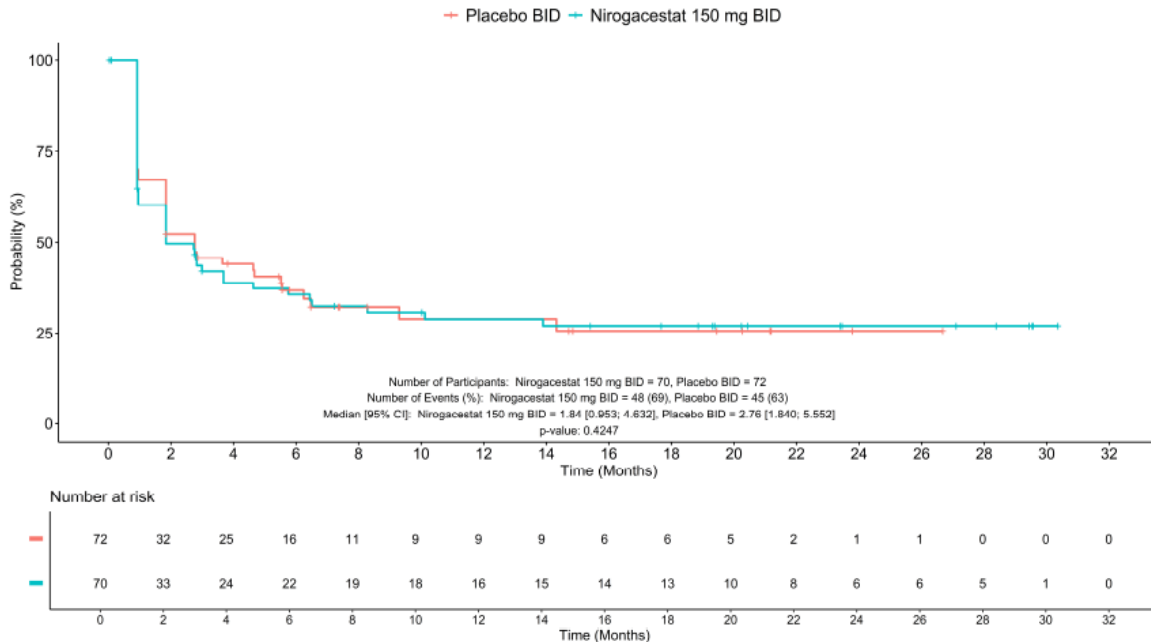


Abbildung 10: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Fatigue“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

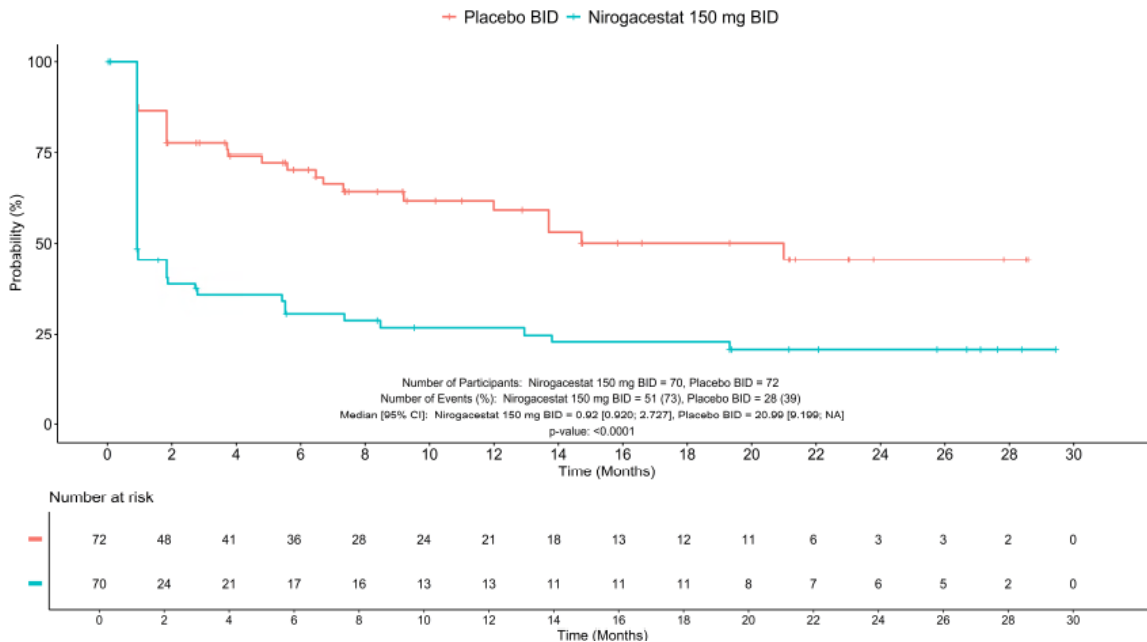


Abbildung 11: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Übelkeit und Erbrechen“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

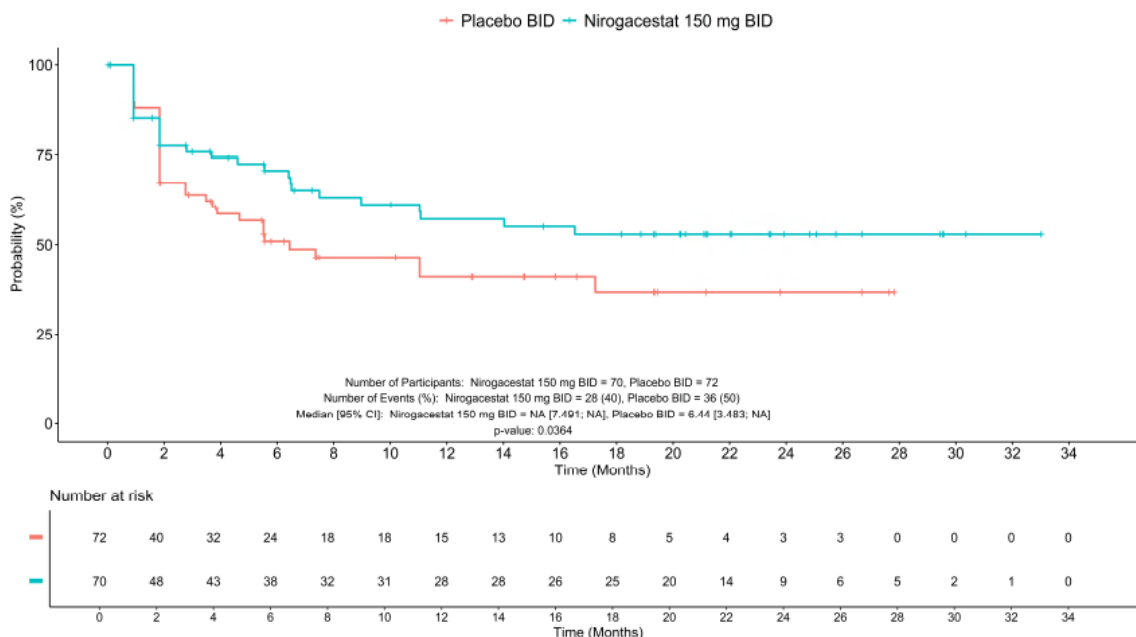


Abbildung 12: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Schmerz“ (Zeit bis zur ersten Verschlechterung um ≥ 10 Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

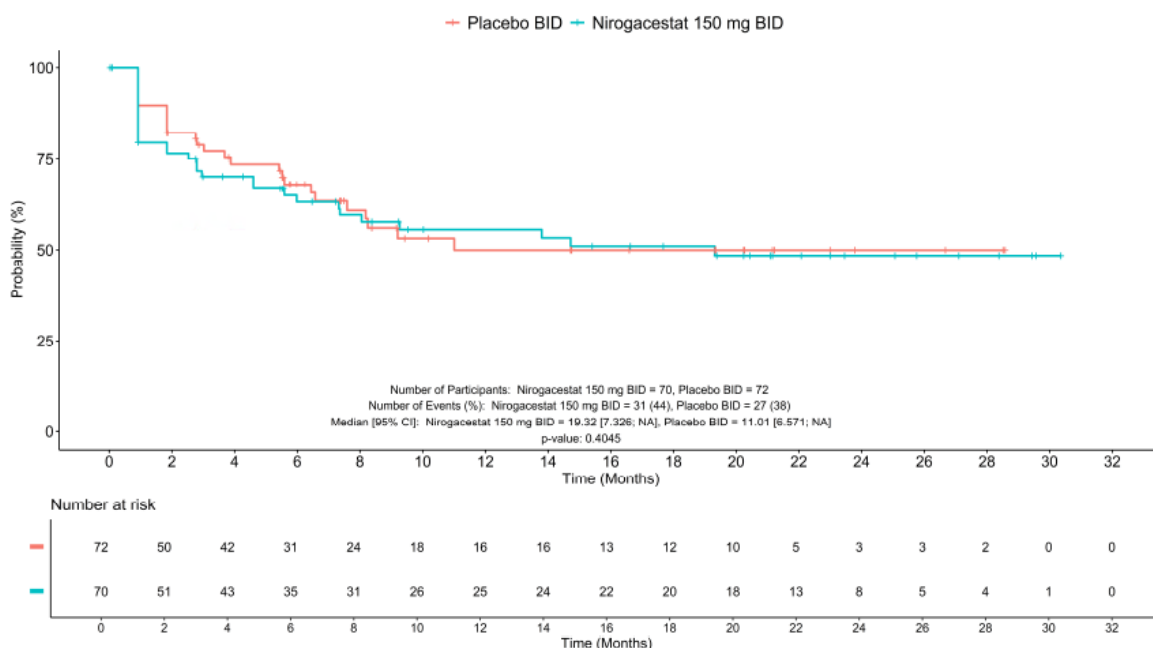


Abbildung 13: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Dyspnoe“ (Zeit bis zur ersten Verschlechterung um ≥ 10 Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

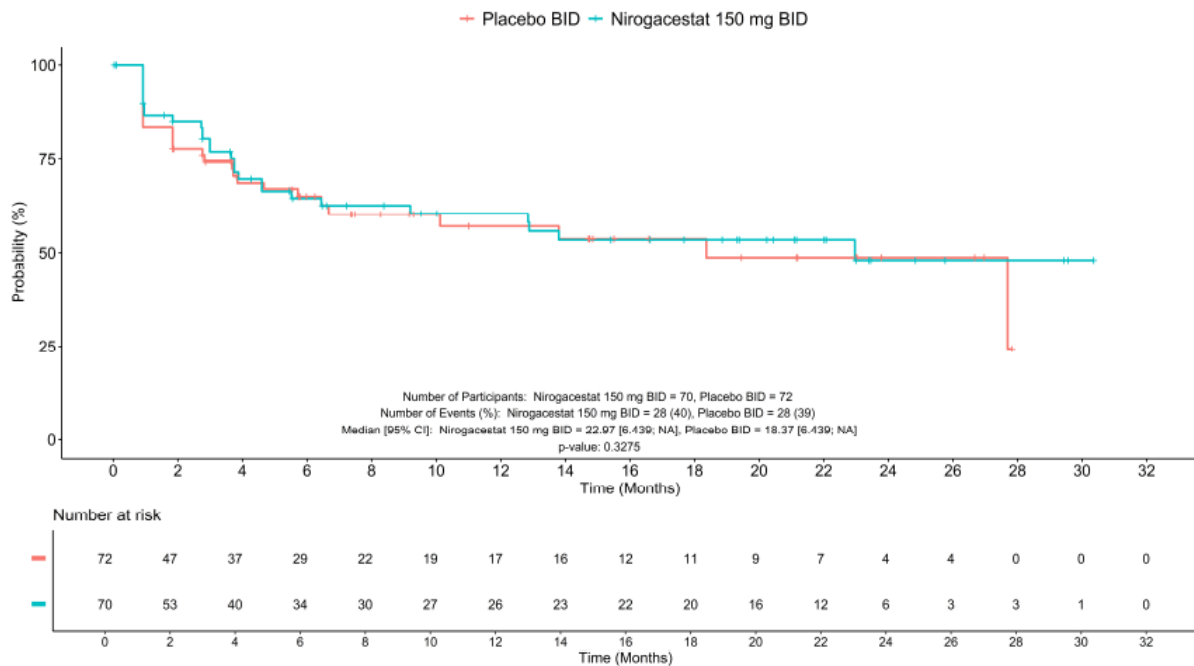


Abbildung 14: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Schlaflosigkeit“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

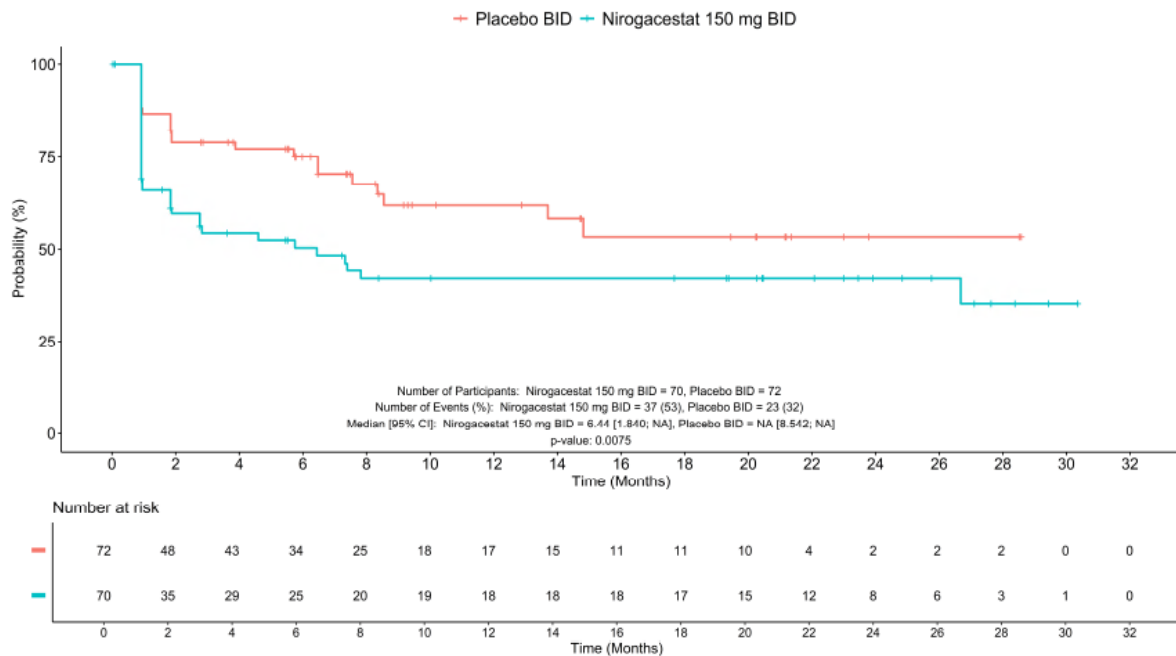


Abbildung 15: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Appetitverlust“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

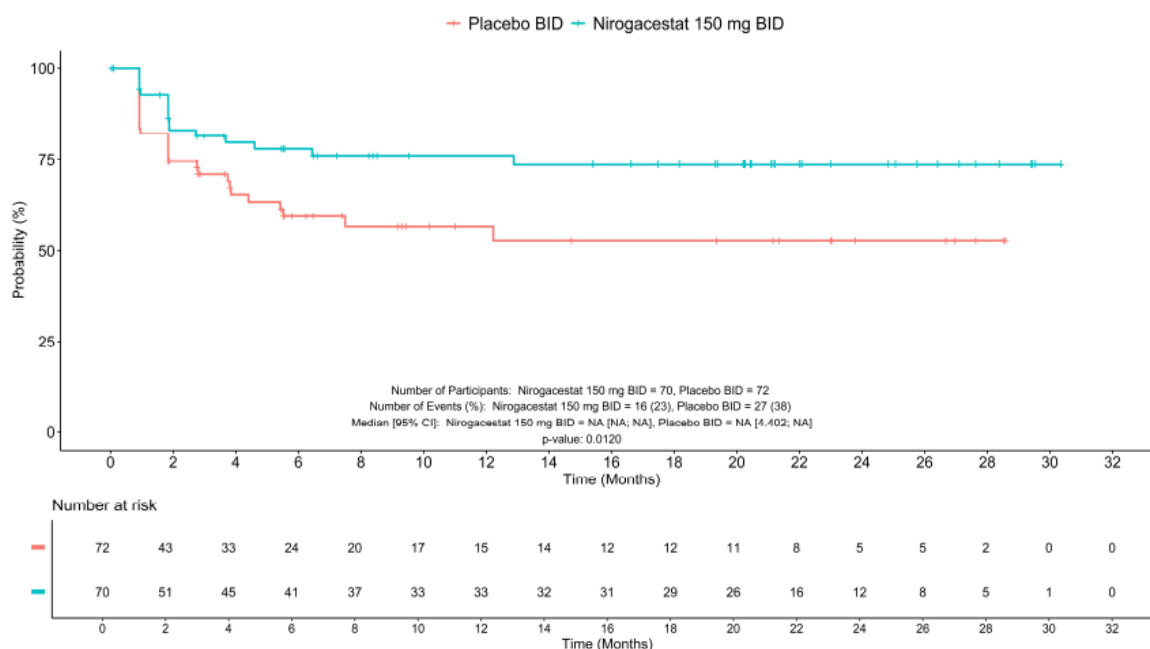


Abbildung 16: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Obstipation“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

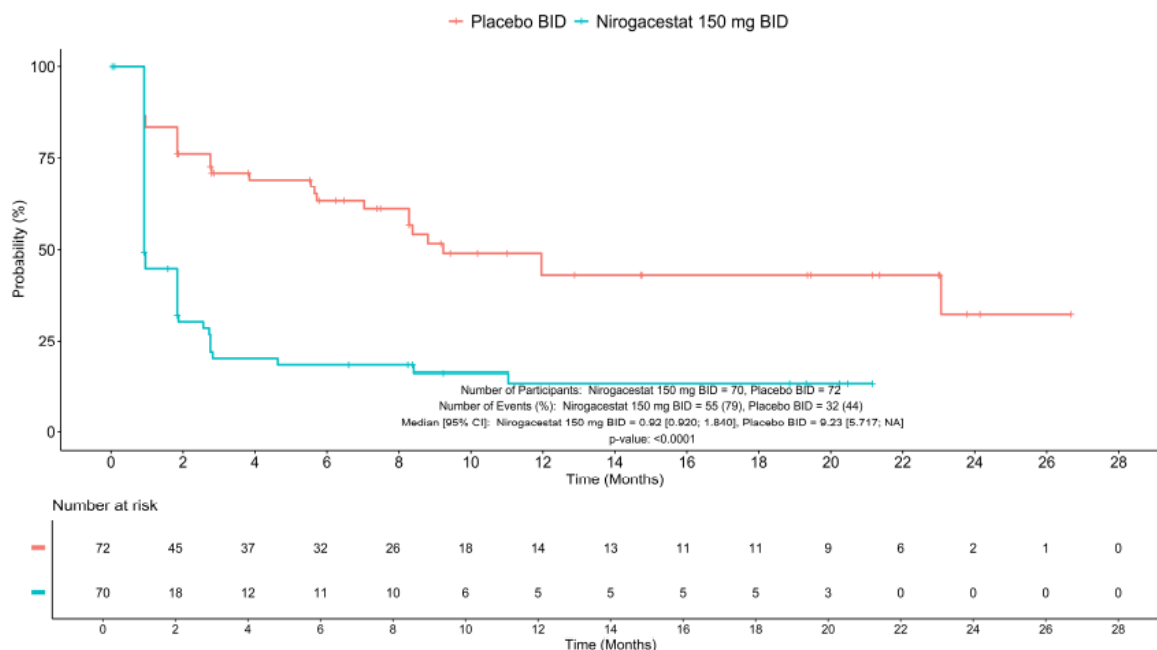


Abbildung 17: Kaplan-Meier-Kurve für EORTC QLQ-C30: Symptomskala „Diarrhö“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

**EORTC QLQ-C30: Funktionskalen und globaler Gesundheitsstatus – Kaplan-Meier-Kurven**

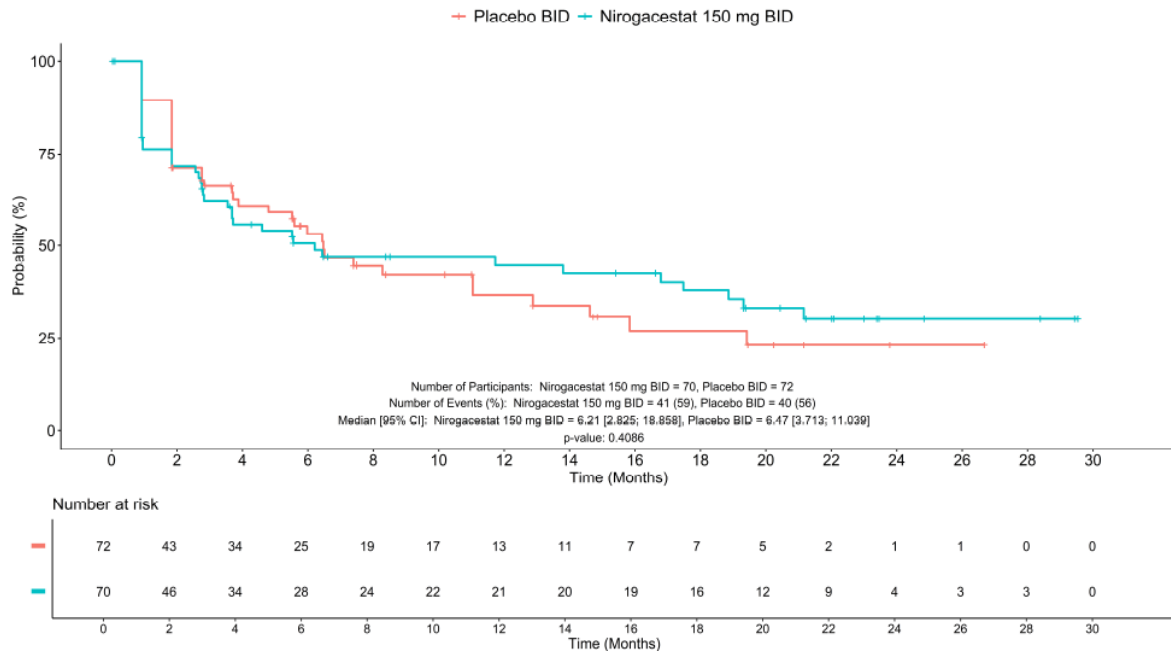


Abbildung 18: Kaplan-Meier-Kurve für EORTC QLQ-C30: Globaler Gesundheitsstatus (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

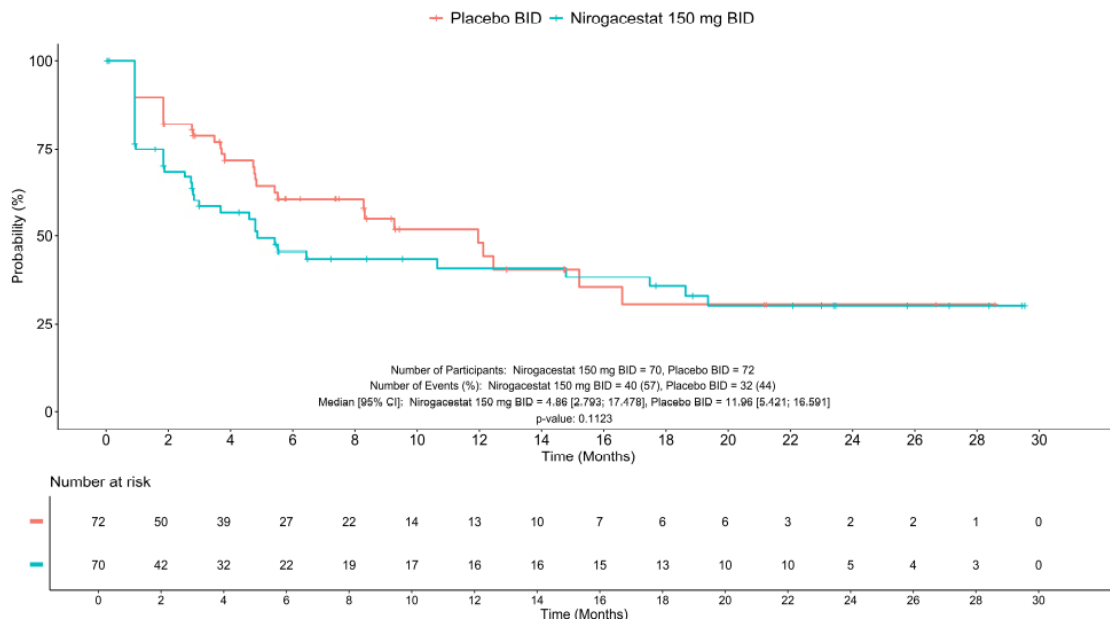


Abbildung 19: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Kognitive Funktion“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

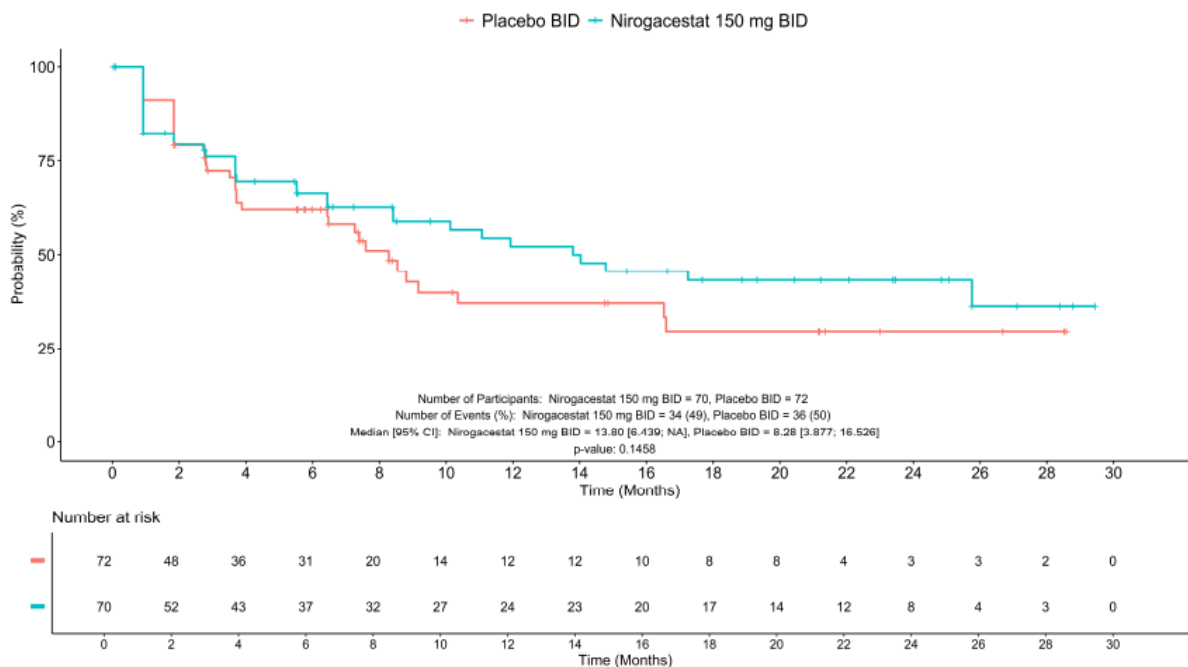


Abbildung 20: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Emotionale Funktion“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

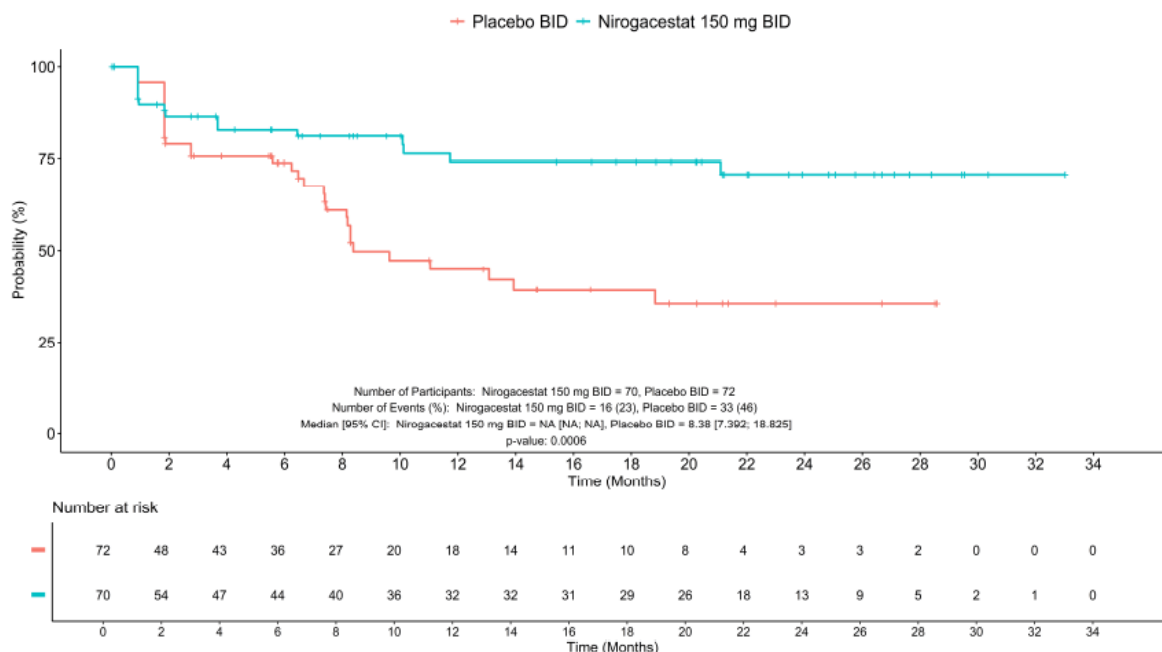


Abbildung 21: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Körperliche Funktion“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

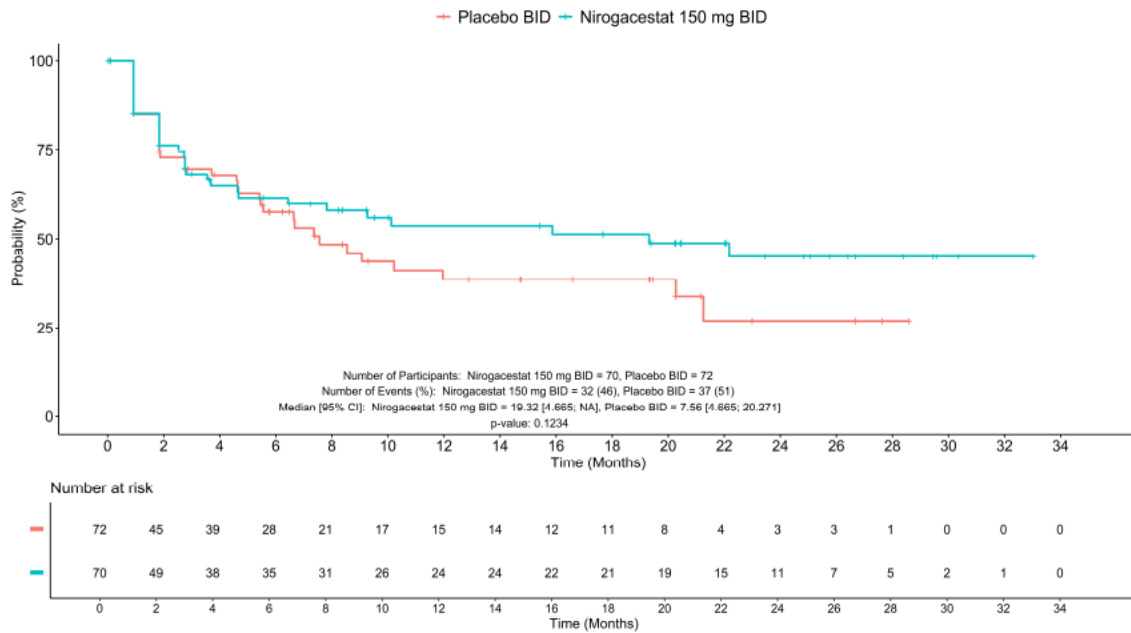


Abbildung 22: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Rollenfunktion“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

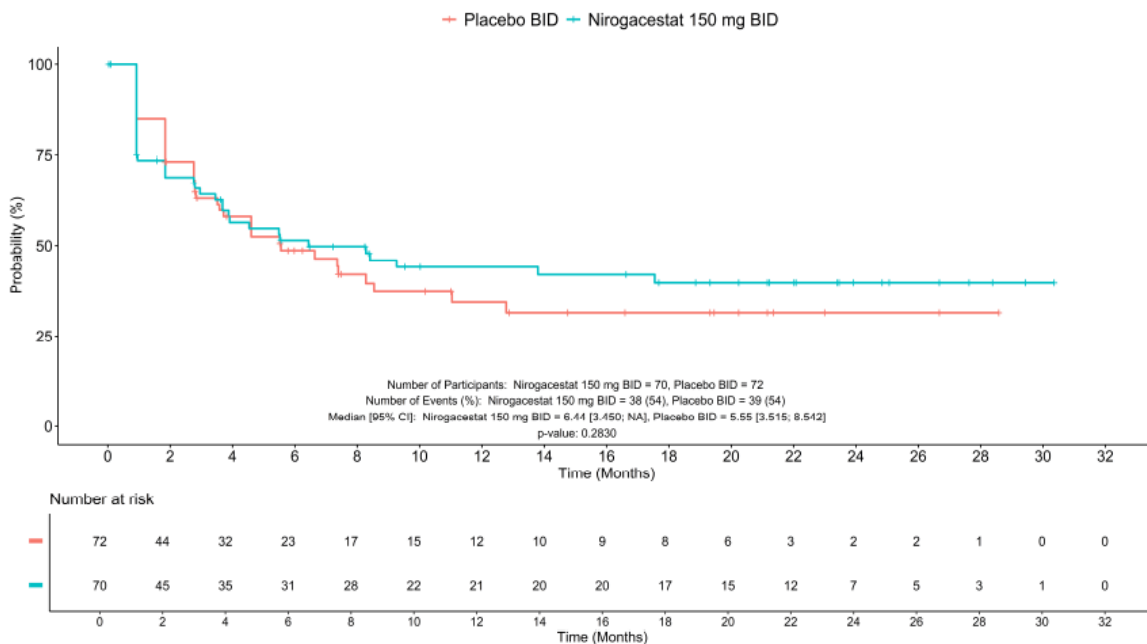


Abbildung 23: Kaplan-Meier-Kurve für EORTC QLQ-C30: Skala „Soziale Funktion“ (Zeit bis zur ersten Verschlechterung um  $\geq 10$  Punkte); Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

**Unerwünschte Ereignisse – Kaplan-Meier-Kurven**

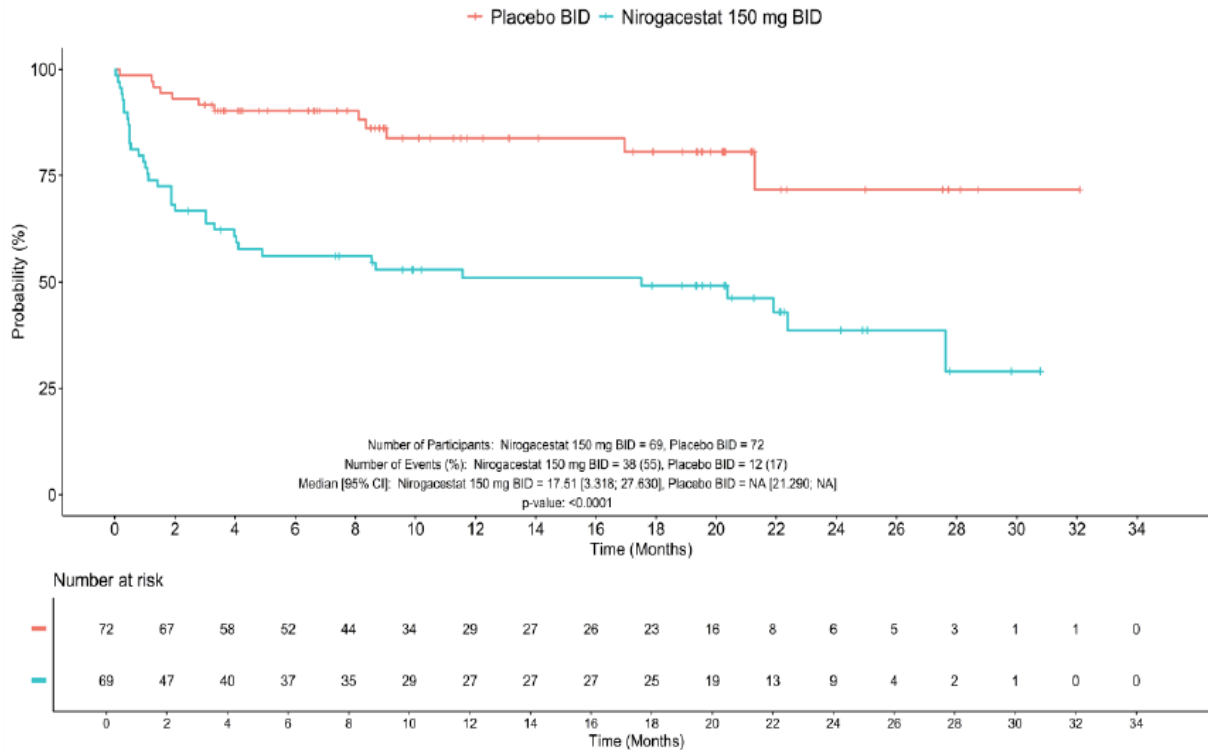


Abbildung 24: Kaplan-Meier-Kurve für schwere UE (CTCAE-Grad  $\geq 3$ ), Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

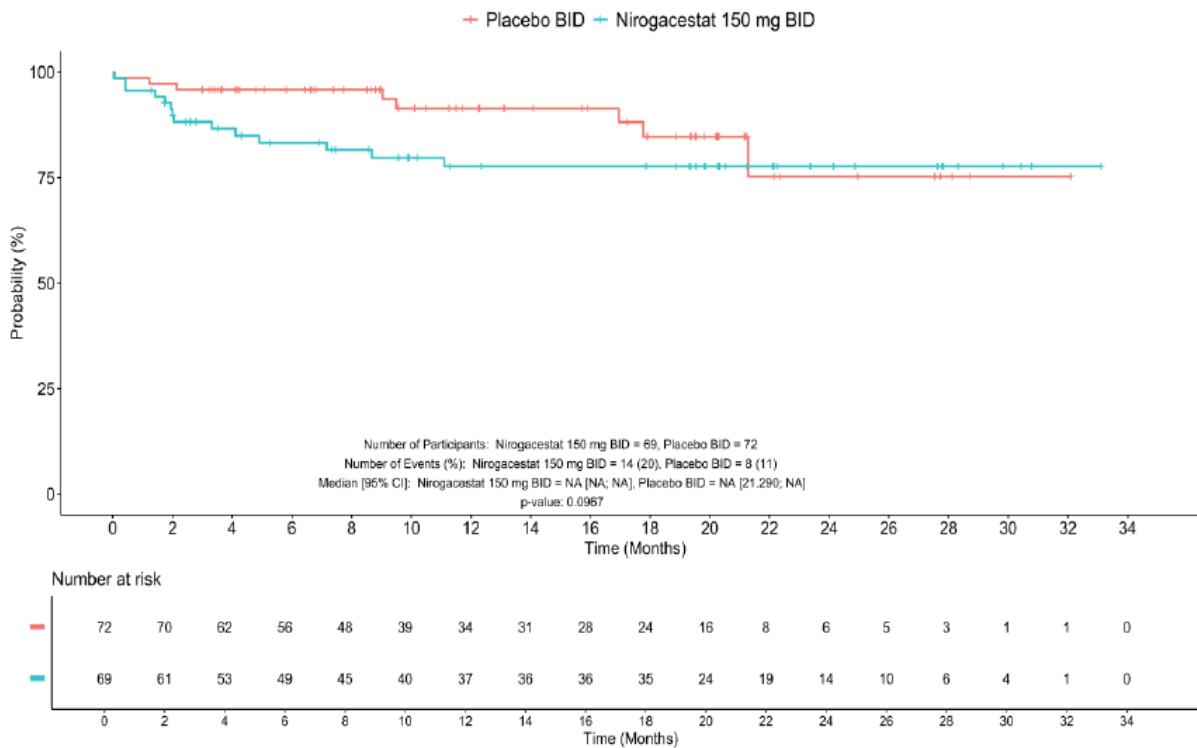


Abbildung 25: Kaplan-Meier-Kurve für SUE, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)

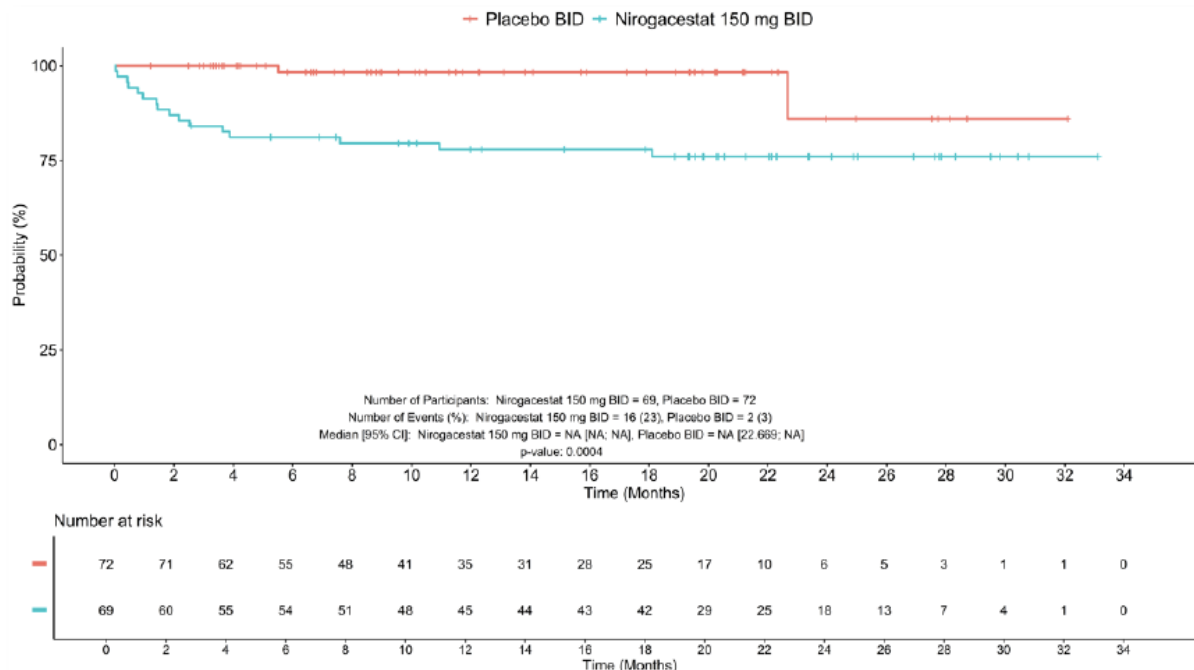


Abbildung 26: Kaplan-Meier-Kurve UE, die zum Abbruch der Studienmedikation führten, Zeit bis zum ersten Auftreten; Studie DeFi (DB-Phase), ITT-Population (Datenschnitt: 07.04.2022)