

Nutzenbewertung



Gemeinsamer
Bundesausschuss

von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V

**Bewertung von Arzneimitteln für seltene Leiden nach § 35a
Absatz 1 Satz 11 i. V. m. Kapitel 5 § 12 Nr. 1 Satz 2 Verfo**

Wirkstoff: Omaveloxolon

Datum der Veröffentlichung: 17. Juni 2024

Inhaltsverzeichnis

Tabellenverzeichnis	3
Abkürzungsverzeichnis	4
Hintergrund	5
1 Fragestellung	6
2 Auswahl und Methodik der relevanten Studien	7
2.1 Studienbasis für die Nutzenbewertung	7
2.2 Beschreibung der eingeschlossenen Studie	7
2.3 Endpunkte	11
2.3.1 Mortalität	12
2.3.2 Morbidität	12
2.3.3 Lebensqualität	20
2.3.4 Sicherheit	21
2.3.5 Erhebungszeitpunkte	22
2.4 Statistische Methoden	23
2.5 Verzerrungspotential auf Studien- und Endpunktebene	24
2.6 Indirekter Vergleich	25
2.6.1 Zusammenfassende Bewertung der Methodik des indirekten Vergleichs	27
3 Ergebnisse der eingeschlossenen Studie	28
3.1 Studiencharakteristika und Studienmedikation	28
3.2 Mortalität	30
3.3 Morbidität	30
3.4 Lebensqualität	31
3.5 Sicherheit	32
4 Diskussion der Methodik und Ergebnisse	35
4.1 Zulassungsstatus und Zulassungspopulation von Omaveloxolon	35
4.2 Design und Methodik der Studie	35
4.3 Mortalität	37
4.4 Morbidität	37
4.5 Lebensqualität	38
4.6 Sicherheit	38
5 Anforderung an eine qualitätsgesicherte Anwendung	39
6 Zusammenfassung der Nutzenbewertung	40
Referenzen	42

Tabellenverzeichnis

Tabelle 1: Übersicht über die Studienbasis	7
Tabelle 2: Charakterisierung der Studie MOXle.....	7
Tabelle 3: Charakterisierung der Intervention Omaveloxolon in der Studie MOXle Teil 2	10
Tabelle 4: Zusammenfassung der Endpunktbewertung der Studie MOXle Teil 2	11
Tabelle 5: Erhebungszeitpunkte der berücksichtigten Endpunkte in der Studie MOXle Teil 2..	22
Tabelle 6: Verzerrungspotential der Studie MOXle Teil 2	24
Tabelle 7: Verzerrungspotential der in der Nutzenbewertung berücksichtigten Endpunkte der Studie MOXle Teil 2	25
Tabelle 8: Allgemeine Angaben; Studie MOXle Teil 2.....	28
Tabelle 9: Charakterisierung der Studienpopulation der Studie MOXle Teil 2; ITT-Population.	28
Tabelle 10: Responderanalysen für eine Verbesserung bzw. Verschlechterung im PGI-C zu Woche 48, Studie MOXle Teil 2, ITT	30
Tabelle 11: Veränderung im ADL zu Woche 48, Studie MOXle Teil 2; ITT-Population	31
Tabelle 12: Responderanalysen für eine Verschlechterung um 15% im SF-36 zu Woche 48; Studie MOXle Teil 2; ITT-Population.....	31
Tabelle 13: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE; Studie MOXle Teil 2, Sicherheitspopulation.....	32
Tabelle 14: UE jeglichen Schweregrads mit Inzidenz ≥ 10 % (bzw. mind. 10 Personen) in einem Studienarm in der Studie MOXle Teil 2; Sicherheitspopulation	33
Tabelle 15: Zusammenfassende Darstellung der Ergebnisse der Studie MOXle Teil 2	40

Abkürzungsverzeichnis

ADL	Activities of Daily Living
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
CGI-C	Clinical Global Impression of Change
EMA	European Medicines Agency
EPAR	European Public Assessment Report
FA	Friedreich-Ataxie
FA-COMS	Friedreich Ataxia – Clinical Outcome Measures Study
FARS	Friedreich Ataxia Rating Scale
FARSn	FARS Neuro
FAS	Full Analysis Set
FI	Fachinformation
GAA	Guanin-Adenin-Adenin
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to Treat
k. A.	keine Angabe
KI	Konfidenzintervall
LS	Least Squares
MCS	Mental Component Summary
MedDRA	Medical Dictionary for Regulatory Activities
mFARS	modified Friedreich Ataxia Rating Scale
MMRM	Mixed Model for Repeated Measurement
MW	Mittelwert
N	Anzahl
9-HPT	9-Hole Peg Test
OLE	Open-Label Extension
PASS	Post-Authorisation Safety Study
PCS	Physical Component Summary
PGI-C	Patient Global Impression of Change
PT	Preferred Term/s
pU	pharmazeutischer Unternehmer
RR	Relatives Risiko
SAP	Statistischer Analyseplan
SF-36	Short-Form-36 Health Survey
SGB	Sozialgesetzbuch
SOC	Systemorganklasse
SUE	Schwerwiegendes unerwünschtes Ereignis
T25-FWT	Timed 25-Foot Walk Test
UE	Unerwünschtes Ereignis
VerfO	Verfahrensordnung des G-BA

Hintergrund

Omaveloxolon 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 Omaveloxolon 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 Omaveloxolon in seiner Sitzung am 11. Juni 2024 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 15. März 2024 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pU in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

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

Die Nutzenbewertung wird am 17. Juni 2024 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1 Fragestellung

Omaveloxolon (Skyclarys®) 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 [23]:

Behandlung der Friedreich-Ataxie bei Erwachsenen und Jugendlichen ab 16 Jahren.

Gemäß Fachinformation beträgt die zugelassene Dosierung 150 mg Omaveloxolon (3 Hartkapseln zu je 50 mg) einmal täglich [23].

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
Studien zum Wirkstoff				
MOXIe (408-C-1402) ¹⁾	Ja	Ja	Ja	-
Studien zu externen Kontrollen				
Friedreich Ataxia – Clinical Outcome Measures Study (FA-COMS) ²⁾	Ja	Supportiv	Nein	Siehe Kapitel 2.6

¹⁾ Zulassungsrelevante Studie gemäß EPAR.

²⁾ Im Rahmen des Zulassungsverfahrens bei der EMA supportiv zu der Zulassungsstudie eingereicht.

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

Zur Nutzenbewertung für Omaveloxolon herangezogene Unterlagen

- Dossier des pU zu Omaveloxolon [1]
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report [6]
- Studienbericht inkl. Studienprotokoll, statistischem Analyseplan (SAP) und Zusatzanalysen der Studie MOXIe (408-C-1402) [24,25,29,30]
- Studienbericht und SAP zum indirekten Vergleich MOXIe und FA-COMS [26,27,28]
- Fachinformation zu Omaveloxolon [23]
- Publikationen zu Omaveloxolon [12,13,14,17]

2.2 Beschreibung der eingeschlossenen Studie

Die Nutzenbewertung von Omaveloxolon im vorliegenden Anwendungsgebiet basiert auf der Zulassungsstudie 408-C-1402 (MOXIe). Die Studie und die Intervention werden in den Tabellen 2-3 charakterisiert.

Tabelle 2: Charakterisierung der Studie MOXIe

Charakteristikum	Beschreibung
Design und Studienablauf	<p>Die MOXIe-Studie ist eine multizentrische, randomisierte, kontrollierte, doppelverblindete Phase-II-Studie bei Patientinnen und Patienten mit Friedreich-Ataxie. Die Studie besteht aus 3 Teilen:</p> <p>Teil 1: Randomisiert-kontrollierte, doppelblinde Dosisfindungsphase zur Bestimmung der Omaveloxolon-Dosis für Teil 2</p> <ul style="list-style-type: none"> • 3:1-Randomisierung in bis zu 9 Kohorten mit jeweiligen Dosierungen von 2,5 mg, 5 mg, 10 mg, 20 mg, 40 mg, 80 mg, 160 mg und 300 mg Omaveloxolon (n = 6) oder Placebo (n = 2) • 12-wöchige Behandlungsphase • 4-wöchige Nachbeobachtung der Sicherheit

Charakteristikum	Beschreibung
	<p>Teil 2: Randomisiert-kontrollierte, doppelblinde Studienphase zur Untersuchung der Wirksamkeit und Sicherheit von Omaveloxolon (150 mg) vs. Placebo</p> <ul style="list-style-type: none"> • 1:1-Randomisierung, stratifiziert nach Pes-cavus-Status¹⁾ (Pes cavus vs. kein Pes cavus) • 48-wöchige, kontrollierte Behandlungsphase (FI-konforme Dosierung) • 4-wöchige Nachbeobachtung der Sicherheit <p>Teil 3: Open-Label-Extensionsphase zur Untersuchung der Langzeitsicherheit und -wirksamkeit</p> <ul style="list-style-type: none"> • Alle Personen erhalten Omaveloxolon in FI-konformer Dosierung. • 144-wöchige, einarmige Behandlungsphase <p>Teil 1 und 2 (RCT-Phasen) umfassen unterschiedliche Patientinnen und Patienten, die nach Abschluss von Teil 1 und Teil 2 und Prüfung der Eignung²⁾ in die einarmige Langzeit-Extensionsphase (OLE) übergehen konnten. Die Patientinnen und Patienten wurden bei Eintritt in die OLE nicht von der Studienbehandlung in Teil 1 oder Teil 2 entblindet. Zum Zeitpunkt der Nutzenbewertung sind Teil 1 und Teil 2 der Studie abgeschlossen; Teil 3 der Studie läuft noch.</p> <p>Zur Beantwortung der Fragestellung für die Nutzenbewertung wird die doppelblinde RCT-Phase (Teil 2) herangezogen.</p>
Population	<p>Wesentliche Einschlusskriterien</p> <ul style="list-style-type: none"> • Genetisch bestätigte Friedreich-Ataxie (FA) • mFARS-Score ≥ 20 und ≤ 80. Der Durchschnitt der beiden mFARS-Scores, die beim Screening und an Tag 1 ermittelt wurden, mussten innerhalb eines zulässigen Bereichs liegen und durften nicht mehr als 4,5 Punkte voneinander abweichen. • Männlich oder weiblich und ≥ 16 Jahre und ≤ 40 Jahre alt • Keine Veränderung des Trainingsprogramms in den letzten 30 Tagen vor dem ersten Studientag und Bereitschaft, dieses Trainingsprogramm während des Studienzeitraums beizubehalten • Patientinnen und Patienten mussten in der Lage sein, den „Maximal Exercise Test“ durchzuführen. • Intakte Nierenfunktion, definiert als geschätzte glomeruläre Filtrationsrate (eGFR) ≥ 60 ml/min/1,73 m² unter Verwendung der 4-Variablen-Formel der Modification of Diet in Renal Disease • Linksventrikuläre Ejektionsfraktion ≥ 40 % (basierend auf einem Echokardiogramm, das beim Screening oder innerhalb von 90 Tagen vor dem Screening durchgeführt wurde) <p>Wesentliche Ausschlusskriterien</p> <ul style="list-style-type: none"> • Unkontrollierter Diabetes (HbA1c $> 11,0$ %) • BNP-Wert > 200 pg/ml • Klinisch signifikante linksseitige Herzerkrankung und/oder eine klinisch signifikante Herzerkrankung in der Vorgeschichte, mit Ausnahme einer leichten bis mittelschweren Kardiomyopathie im Zusammenhang mit FA, einschließlich, aber nicht beschränkt auf eine der folgenden Erkrankungen: <ul style="list-style-type: none"> ○ klinisch signifikante angeborene oder erworbene Herzklappenerkrankung ○ Herzbeutelverengung (basierend auf einem Echokardiogramm, das zum Screening oder innerhalb von 90 Tagen vor dem Screening durchgeführt wurde) ○ restriktive oder kongestive Kardiomyopathie (basierend auf einem Echokardiogramm, das zum Screening oder innerhalb von 90 Tagen vor dem Screening durchgeführt wurde)

Charakteristikum	Beschreibung
	<ul style="list-style-type: none"> ○ symptomatische Koronarerkrankung (früherer Myokardinfarkt, perkutane Koronarintervention, koronare Bypassoperation oder Angina pectoris) ○ Krankenhausaufenthalt wegen Herzinsuffizienz in den letzten 5 Jahren ○ Herzinsuffizienz, definiert als New York Heart Association Klasse > 2 ○ Vorgeschichte von Vorhofflimmern ○ instabile Herzrhythmusstörungen in der Vorgeschichte ● Klinisch signifikante Anomalien in der klinischen Hämatologie oder Biochemie, einschließlich, aber nicht beschränkt auf Erhöhungen der AST- oder ALT-Werte um mehr als das 1,5 x ULN (Werte über diesem Grenzwert waren zulässig, wenn sie auf eine Muskelverletzung zurückzuführen waren.) ● Patientinnen und Patienten, die innerhalb von 7 Tagen vor dem ersten Studientag eines der folgenden Medikamente eingenommen hatten oder die Einnahme eines dieser Medikamente während der Studienteilnahme planten: <ul style="list-style-type: none"> ○ empfindliche Substrate für Cytochrom P450 2C8 oder 3A4 (z. B.: Repaglinid, Midazolam, Sildenafil) ○ mäßige oder starke Hemmstoffe oder Induktoren von Cytochrom P450 3A4 (z. B. Carbamazepin, Phenytoin, Ciprofloxacin, Grapefruitsaft) ○ Substrate für den p-Glykoprotein-Transporter (z. B. Ambrisentan, Digoxin) ● Klinisch signifikante Lebererkrankung in der Vorgeschichte (z. B. Fibrose, Zirrhose, Hepatitis) oder beim Screening klinisch relevante Abweichungen in den Labortests, einschließlich einer der folgenden Punkte: <ul style="list-style-type: none"> ○ ALT und/oder AST > 1,5 x ULN ○ Bilirubin > 1,2 x ULN ○ alkalische Phosphatase > 2 x ULN ○ Albumin < Lower Limit of Normal (LLN) ● Patientinnen und Patienten, die innerhalb von 14 Tagen vor dem ersten Studientag antioxidative Präparate, einschließlich, aber nicht beschränkt auf Idebenon, Coenzym Q10, Nikotinamid und Vitamin E, eingenommen hatten, welche über der empfohlenen Tagesdosis lagen, oder die Einnahme eines dieser Präparate während der Studienteilnahme planten ● Dokumentierte mitochondriale Atmungskettenerkrankung in der Vorgeschichte ● Innerhalb der letzten 5 Jahre thromboembolische Ereignisse in der Vorgeschichte ● 30 Tage vor dem 1. Studientag Einnahme einer gerinnungshemmenden Therapie ● Patientinnen und Patienten, bei denen während der Studie eine chirurgische Behandlung wegen Skoliose oder Fußdeformität geplant war
Intervention und Zahl der Patientinnen und Patienten	Gescreent gesamt: N = 156 Randomisiert/Eingeschlossen: N = 103 Intervention: N = 51 Kontrolle: N = 52
Ort und Zeitraum der Durchführung	Studienzentren 11 Studienzentren (USA: N = 7; Australien: N = 1; Österreich: N = 1; Italien: N = 1; Vereinigtes Königreich: N = 1) Studienzeitraum <ul style="list-style-type: none"> ● Erste Person erste Visite: 20. Oktober 2017 ● Letzte Person letzte Visite: 31. Oktober 2019

Charakteristikum	Beschreibung
Endpunkte gemäß Studienprotokoll	<p>Primäre Endpunkte</p> <ul style="list-style-type: none"> • modified Friedreich Ataxia Rating Scale (mFARS)-Score • Sicherheit <p>Sekundäre Endpunkte</p> <ul style="list-style-type: none"> • Maximale Belastbarkeit, Peak Work Maximal Exercise Test • Patient Global Impression of Change (PGI-C) • Clinical Global Impression of Change (CGI-C) <p>Explorative Endpunkte</p> <ul style="list-style-type: none"> • Short Form (SF)-36 • 9-Hole Peg Test (9-HPT) • Timed 25-Foot Walk Test (T25-FWT) • Einschränkung bei Alltagsaktivitäten (ADL) • Häufigkeit von Stürzen

¹⁾ Pes cavus ist definiert als eine muskuloskelettale Fußdeformität, die durch ein hohes Fußgewölbe gekennzeichnet ist (= Hohlfuß). Die Subpopulation mit Pes cavus war auf 20 % der gesamten randomisierten Population begrenzt.

²⁾ Prüfung auf Eignung zur Teilnahme an OLE: Die Studienteilnehmenden mussten eine 12-wöchige Behandlung in Teil 1 oder eine 48-wöchige Behandlung mit Omeveloxolon in Teil 2 abgeschlossen haben, durften keine Protokollabweichungen aufweisen und die Einschluss- und Ausschlusskriterien der OLE erfüllen.

Abkürzungen: ADL: Activities of Daily Living; BNP: B-Typ natriuretisches Peptid; ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; CGI-C: Clinical Global Impression of Change; FA: Friedreich-Ataxie; FI: Fachinformation; LLN: Unterer Normbereich; mFARS: modified Friedreich Ataxia Rating Scale; 9-HPT: 9-Hole Peg Test; OLE: Open-Label Extension; PGI-C: Patient Global Impression of Change; SF-36: Short-Form-36 Health Survey; T25-FWT: Timed 25-Foot Walk Test; ULN: Oberer Normbereich.

Protokolländerungen

Es wurden 8 Änderungen des Originalprotokolls vom 23. Juli 2014 bis zur Veröffentlichung des Studienberichts vom 05. November 2020 vorgenommen. Alle Patientinnen und Patienten in der Studie MOXle Teil 2 wurden im Rahmen der Version 9 (Amendment 8) des Studienprotokolls vom 17. August 2017 in die Studie aufgenommen. Eine relevante Änderung dieses Amendments war, dass die Dauer der Studie MOXle Teil 2 von 24 auf 48 Wochen verlängert wurde.

Charakterisierung der Intervention

Tabelle 3: Charakterisierung der Intervention Omeveloxolon in der Studie MOXle Teil 2

Intervention	Kontrolle
<p>Omeveloxolon</p> <ul style="list-style-type: none"> • Dosierung: 150 mg Omeveloxolon (3 Hartkapseln zu je 50 mg) einmal täglich; peroral <p>Studienteilnehmende mussten das Studienmedikament dauerhaft absetzen, wenn eine der folgenden Bedingungen eintrat:</p> <ul style="list-style-type: none"> • ALT oder AST > 8 x ULN • ALT oder AST > 5 x ULN für mehr als 2 Wochen • ALT oder AST > 3 x ULN und TBL > 2 x ULN • ALT oder AST > 3 x ULN mit dem Auftreten von Müdigkeit, Übelkeit, Erbrechen, Schmerzen im rechten oberen Quadranten, Fieber, Hautausschlag und/oder Eosinophilie (> 5 %) 	<p>Placebo</p> <ul style="list-style-type: none"> • Wie bei Intervention, ohne den aktiven Wirkstoff <p>Regeln zum Abbruch der Studienmedikation:</p> <ul style="list-style-type: none"> • Wie bei Intervention

Intervention	Kontrolle
Nicht erlaubte Begleitmedikation <ul style="list-style-type: none"> • Gerinnungshemmende Medikamente, mit Ausnahme von täglichem „Baby-Aspirin“ (bis zu 81 mg) • Antioxidative Präparate, einschließlich, aber nicht beschränkt auf Idebenon, Coenzym Q10, Nikotinamid und Dosierungen, die die empfohlene Tagesdosis an Vitamin E überschreiten • Empfindliche Substrate für Cytochrom P450 2C8 oder 3A4 (z. B. Repaglinid, Midazolam, Sildenafil) • Mäßige oder starke Hemmstoffe oder Induktoren von Cytochrom P450 3A4 (z. B. Carbamazepin, Phenytoin, Ciprofloxacin, Grapefruitsaft) • Substrate für den p-Glykoprotein-Transporter (z. B. Ambrisentan, Digoxin) • Jegliche interventionelle Therapie zur Behandlung der Friedreich-Ataxie-Krankheit, mit Ausnahme eines konstanten Übungsprogramms • Antispastika • Pflanzliche Präparate, Vitamine, Mineralien oder rezeptfreie Medikamente 	
Erlaubte Begleitmedikation <ul style="list-style-type: none"> • Antibiotika • Tägliches Multivitamin • Schmerzmittel • Andere Medikamente zur Behandlung gleichzeitiger Erkrankungen • Orale, implantierbare oder injizierbare Verhütungsmittel 	

Abkürzungen: ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; ULN: Oberer Normbereich.

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 4 nicht gelistet sind, wurden weder vom pU noch im Rahmen der Nutzenbewertung als patientenrelevant bzw. bewertungsrelevant eingestuft. Tabelle 4 stellt das Ergebnis dieser Bewertung zusammenfassend dar.

Tabelle 4: Zusammenfassung der Endpunktbewertung der Studie MOXie Teil 2

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
Todesfälle	Mortalität	Ja	Ja
mFARS-Score ¹⁾	Morbidität	Ja	Nein
Maximale Belastbarkeit, Peak Work Maximal Exercise Test		Ja	Nein
PGI-C		Ja	Ja
CGI-C		Ja	Nein
9-HPT		Ja	Ja
T25-FWT		Ja	Ja
ADL		Ja	Ja
Häufigkeit von Stürzen		Ja	Nein

Endpunkt	Kategorie	Berücksichtigung im Dossier des pU	Berücksichtigung in der Nutzenbewertung
SF-36	Lebensqualität	Ja	Ja
Unerwünschte Ereignisse ¹⁾	Sicherheit	Ja	Ja

¹⁾ Primärer Endpunkt.

Abkürzungen: ADL: Activities of Daily Living; CGI-C: Clinical Global Impression of Change; mFARS: modified Friedreich Ataxia Rating Scale; 9-HPT: 9-Hole Peg Test; PGI-C: Patient Global Impression of Change; SF-36: Short-Form-36 Health Survey; T25-FWT: Timed 25-Foot Walk Test.

2.3.1 Mortalität

Todesfälle

Der Endpunkt „Todesfälle“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Todesfälle werden im Rahmen der Beobachtung der Sicherheitsendpunkte von Baseline (erste Einnahme der Studienmedikation) bis zum Behandlungsende bzw. bis 30 Tage nach der letzten Einnahme der Studienmedikation bei vorzeitigem Behandlungsende erhoben.

Bewertung

Die Operationalisierung ist nachvollziehbar.

Patientenrelevanz und Validität

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

2.3.2 Morbidität

mFARS – modified Friedreich Ataxia Rating Scale

Der Endpunkt „mFARS“ zur Erhebung der Krankheitschwere wird in der Nutzenbewertung aufgrund nicht nachvollziehbarer Operationalisierung und unklarer Validität nicht berücksichtigt.

Operationalisierung

Beschreibung

Die modifizierte Friedreich Ataxia Rating Scale (mFARS) misst mittels Funktionstests die körperlichen Funktionen von Patientinnen und Patienten mit Friedreich-Ataxie (FA) in 4 Domänen: bulbäre Funktion, Koordination der oberen Extremitäten, Koordination der unteren Extremitäten und aufrechte Stabilität. Die mFARS basiert auf den Instrumenten FARS und FARSn (Neuro). Die FARS wurde 2005 entwickelt. Für die FARSn wurden einige Items der FARS gestrichen und 2 weitere Items hinzugefügt (maximal zu erreichende Punktzahl: 125). Für die mFARS wurden Items der FARSn gestrichen, die nicht direkt funktionale Fähigkeiten bewerten (Domäne zum peripheren Nervensystem sowie 2 bulbäre Komponenten (A1 und A2), siehe Abbildung 1). Zur Berechnung des Gesamtscores der mFARS werden die einzelnen Werte addiert (maximal zu erreichende Punktzahl: 93). Ein höherer Score indiziert eine stärkere körperliche Beeinträchtigung. In der Studie MOXie Teil 2 umfasst der Gesamtscore 99 Punkte statt 93 Punkten. Ausgewertet wurden die 4 Domänen bulbäre Funktion, Koordination der oberen Extremitäten, Koordination der unteren Extremitäten und aufrechte Stabilität, einschließlich der 2 bulbären Komponenten A1 und A2. Die Domäne peripheres Nervensystem wird bei der Bildung des Gesamtscores nicht berücksichtigt.

		FARSn (N = 125)	mFARS (N = 93)
A1** (3) Facial atrophy A2** (3) Tongue atrophy	A3 (2) Cough A4 (3) Speech	Bulbar (11)	Bulbar – (5)
B1 (3+3) Finger-finger B2 (4+4) Nose-finger B3 (4+4) Dysmetria	B4 (3+3) Rapid movements B5 (4+4) Finger taps	Upper limb coordination (36)	Upper limb coordination (36)
C1 (4+4) Heel-shin slide C2 (4+4) Heel-shin tap		Lower limb coordination (16)	Lower limb coordination (16)
D1 (2+2) Muscle atrophy D2 (5+5) Musc. weakness D3 (2+2) Vibratory sense	D4 (2+2) Position sense D5 (2+2) Deep tendon reflexes	Peripheral nervous system (26)	
E1 (4) Sitting position E2A (4) Stance feet apart E2B (4) With eyes closed E3A (4) Stance, feet together. E3B (4) With eyes closed	E4 (4) Tandem stance E5 (4) Stance, dom. foot E6 (3) Tandem walk E7 (5) Gait	Upright stability (36)	Upright stability (36)

Maximum score/subscale/item scores are shown in brackets. Items in subscales B, C, and D are conducted separately on lateral sides; ** items A1 and A2 are excluded in the mFARS examination. FARS = Friedreich Ataxia Rating Scale; mFARS = modified FARS.

Abbildung 1: Messmodell der neurologischen Untersuchungen des FARSn und des mFARS [31]

In der Studie MOXle Teil 2 wurde der umfangreichere FARSn erhoben. Die Erhebung erfolgte durch einen Neurologen / eine Neurologin; wenn möglich sollte zu allen Erhebungszeitpunkten dieselbe Person den Test mit dem Patienten / der Patientin durchführen. Zusätzlich gab es eine Anleitung zur Testdurchführung (Studienmanual). Patienten/Patientinnen, die den Test noch nicht kannten, konnten einen Übungstest während des Screenings machen. Mit Ausnahme des Übungstests sollte der Test immer nach dem maximalen Belastungstest durchgeführt werden.

Neben der mittleren Veränderung sollte der Anteil an Personen mit einer Verbesserung bzw. Verschlechterung ($\leq -1,9$ bzw. $\geq 1,9$) berechnet werden. Die Auswertungen zum ausführlicheren FARSn sollten laut Protokoll explorativ sein.

Wenn bei den Domänen bulbäre Funktion, Koordination der oberen Extremitäten, Koordination der unteren Extremitäten und peripheres Nervensystem eine Antwort fehlte, sollte der Wert der letzten Erhebung übernommen werden. Bei der Domäne aufrechte Stabilität sollte bei den Fragen 1 bis 5 der Mittelwert der aktuellen Antworten gebildet werden. Nur wenn keine Antwort vorliegt sowie bei Frage 6 und 7 soll der Wert der letzten Erhebung übertragen werden.

Der mFARS-Baseline-Wert diene als Einschlusskriterium für die Studienteilnahme. Für die Bildung des Baseline-Wertes wurde der Mittelwert des mFARS vom Screening und von Tag 1 berechnet. Die beiden Werte sollten sich maximal um 4,5 Punkte unterscheiden. Wenn die Werte die zulässige Differenz von 4,5 Punkten um mindestens das Zweifache überschreiten (d. h. mindestens 9 Punkte Unterschied), wird der mFARS-Wert des Screenings als ungültig betrachtet und es wurde lediglich der Wert von Tag 1 als Baseline-Wert verwendet.

Bewertung

Die Operationalisierung ist nur teilweise nachvollziehbar. Obwohl die mFARS als primärer Endpunkt definiert ist, ist unklar, warum der umfangreichere FARSn erhoben wurde. Darüber hinaus erschließt sich nicht, warum für die Berechnung des mFARS die beiden Items Gesichtsatrophy (A1) und Zungenatrophy (A2) verwendet wurden, die von der mFARS nicht umfasst sind. Somit liegen Auswertungen für eine Version des Fragebogens mit einer maximal

zu erreichenden Punktzahl von 99 anstatt der validierten mFARS mit 93 Punkten vor. Laut pU liegt eine klinisch relevante Veränderung (MID) bei 1,9 Punkten pro Jahr. Die erstmals im Studienbericht genannte MID kann anhand der Referenzpublikation [21] nicht nachvollzogen werden. Für die vorliegende Nutzenbewertung wird eine Relevanzschwelle von 15 % der Skalenspannweite als adäquat angesehen.

Patientenrelevanz

Der Endpunkt „mFARS“ wird als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) angesehen.

Validität

Es liegen keine Hinweise in der Literatur zur Existenz der in der MOXle-Studie eingesetzten Version der mFARS mit 99 Punkten vor. Die folgende Bewertung bezieht sich auf die Version des mFARS (93 Punkte). Es wurden 2 Validierungsstudien identifiziert, an denen 1011 und 172 Personen mit FA teilgenommen haben.

Die Reliabilität und Validität des mFARS (93 Punkte) konnten weitgehend bestätigt werden, wobei die Domäne bulbäre Funktion keine überzeugenden Werte bezüglich der internen Konsistenz aufwies und zum Teil Decken- und Bodeneffekte vorlagen [31,32]. An der Entwicklung des Fragebogens waren keine Betroffenen beteiligt, so dass die Inhaltsvalidität nicht eingeschätzt werden kann. Angaben zur Änderungssensitivität liegen nicht vor.

Da der in der Studie MOXle Teil 2 eingesetzte Fragebogen mit maximal zu erreichender Punktzahl von 99 nicht der Validierung zugrundeliegenden und etablierten Version des Fragebogens mit einer maximal zu erreichenden Punktzahl von 93 entspricht, wird der Endpunkt nicht berücksichtigt.

Maximale Belastbarkeit, Peak Work Maximal Exercise Test

Der Endpunkt „Maximale Belastbarkeit“ mittels Peak Work Maximal Exercise Test wird in der Nutzenbewertung aufgrund nicht-gegebener Validität nicht berücksichtigt.

Operationalisierung

Beschreibung

Die maximale Belastbarkeit der Patientinnen und Patienten wurde mithilfe von Fahrradergometrie auf einem stationären Liegerad erhoben. Die Untersuchung sollte innerhalb der Studienvisite als erster Funktionstest durchgeführt werden.

Es wurde u. a. die maximale Belastung (W/kg), die Herzfrequenz und der systolische Blutdruck erhoben. Im Dossier wurde die maximale Belastung (W/kg) dargestellt. Neben der mittleren Veränderung soll der Anteil an Personen mit einer Verbesserung berechnet werden. Die Responderanalyse wurde im Studienbericht und in Modul 4 nicht dargestellt.

Bewertung

Die Operationalisierung ist weitgehend nachvollziehbar. Es ist unklar, ob sich die Methode des Fahrradfahrens bei Personen mit FA zur Erhebung der kardinalen Funktionen eignet, da das Bewegen der Pedale aufgrund der Ataxie ungenau, verlangsamt und eingeschränkt sein kann. Ein niedriger Wert ist daher nicht nur ein Hinweis auf ein kardiologisches Symptom, sondern ebenfalls auf die motorisch-koordinativen Einschränkungen der unteren Extremitäten.

Patientenrelevanz und Validität

Der pU beschreibt, dass die Prüfung der maximalen Belastbarkeit der frühzeitigen Erkennung von Kardiomyopathien dient. Kardiomyopathien sind als patientenrelevante Symptomatik im vorliegenden Anwendungsgebiet anzusehen, die Patientenrelevanz der maximalen Belastbarkeit ist nicht gegeben. Eine Surrogatvalidierung zum Endpunkt wurde nicht

vorgelegt. Es ist zudem weder beschrieben noch belegt, ab welchem Schwellenwert sich eine Kardiomyopathie erkennen ließe. Eine inhaltlich sinnvolle und verständliche Interpretation von Veränderungen dieses Endpunktes erscheint darüber hinaus schwierig.

PGI-C – Patient Global Impression of Change

Der Endpunkt „PGI-C“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

In der Studie MOXle Teil 2 wurde mithilfe des PGI-C der subjektive Eindruck einer Person zur Veränderung des klinischen Gesundheitszustands im Vergleich zum Behandlungsbeginn erfasst. Der PGI-C bezieht sich auf das patientenberichtete Therapieansprechen und wird anhand einer 7-Punkte-Skala von „sehr verbessert“ (= 1), „stark verbessert“ (= 2), „ein wenig verbessert“ (= 3), „unverändert“ (= 4), „wenig schlechter“ (= 5), „viel schlechter“ (= 6) und „sehr viel schlechter“ (= 7) beantwortet.

Die Patientinnen und Patienten füllten den PGI-C im Anschluss an die Erhebung der neurologischen FARS oder des Clinical Global Impression of Change (CGI-C) aus.

Laut Protokoll wurden für die Auswertung der Personen mit einer Verbesserung Personen als Responder gewertet, wenn sie sich im PGI-C „sehr verbessert“ oder „stark verbessert“ haben. Im Modul 4 und im Studienbericht hingegen wurden Personen mit einer Verbesserung als Responder gewertet, wenn sie < 4 Punkte im PGI-C aufwiesen; Personen mit unverändertem Gesundheitszustand bei 4 Punkten und Personen mit einer Verschlechterung im Gesundheitszustand bei > 4 Punkten. Fehlende Werte wurden nicht imputiert.

Bewertung

Die Operationalisierung ist nachvollziehbar. Die Post-hoc-Auswertung, bei der jegliche Verbesserung (< 4 Punkte) als Responder gewertet wird, wird als sachgerecht bewertet.

Patientenrelevanz

Der Endpunkt wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittel-Nutzenbewertungsverordnung (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 Symptomvariabilität besonders anfällig für diese Form der Verzerrung, den Recall Bias [10,19,20]. Diese Faktoren scheinen relevant für die vorliegende Erhebung zu sein, bei der für einen 48-wöchigen Zeitraum im variablen Erkrankungsbild von Friedreich-Ataxie eine Einschätzung getroffen wird. Trotz dieser möglichen Einschränkung wird der PGI-C für die vorliegende Nutzenbewertung herangezogen.

CGI-C – Clinical Global Impression of Change

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

Operationalisierung

Beschreibung

In der Studie MOXle Teil 2 wurde mithilfe des CGI-C, einer Fremdbeurteilungsskala zur Bewertung durch das Prüfpersonal, die globale Veränderung des klinischen Gesamteindrucks des Studienteilnehmenden im Vergleich zum Studienbeginn erfasst. Die Bewertung des Schweregrads im Vergleich zum Studienbeginn (vor Gabe der Studienmedikation) erfolgt auf einer 7-Punkteskala „sehr starke Verbesserung“ (= 1), „starke Verbesserung“ (= 2), „geringe Verbesserung“ (= 3), „keine Veränderung“ (= 4), „geringe Verschlechterung“ (= 5), „starke Verschlechterung“ (= 6), „sehr starke Verschlechterung“ (= 7).

Das Prüfpersonal kann frühere Ergebnisse von Studienbewertungen zur klinischen Beurteilung des CGI-C berücksichtigen. Das Prüfpersonal füllte diese Bewertung im Anschluss an die Erhebung der neurologischen FARS oder des PGI-C aus.

Laut Protokoll wurden für die Auswertung der Personen mit einer Verbesserung Personen als Responder gewertet, wenn sie sich im CGI-C „sehr verbessert“ oder „stark verbessert“ haben. Im Modul 4 und im Studienbericht hingegen wurden Personen mit einer Verbesserung als Responder gewertet, wenn sie < 4 Punkte im CGI-C aufwiesen; Personen mit unverändertem Gesundheitszustand bei 4 Punkten und Personen mit einer Verschlechterung im Gesundheitszustand bei > 4 Punkten. Fehlende Werte wurden nicht imputiert.

Bewertung

Die Operationalisierung ist nachvollziehbar. Die Post-hoc-Auswertung, bei der jegliche Verbesserung (< 4 Punkte) als Responder gewertet wird, wird als sachgerecht bewertet.

Patientenrelevanz

Die Einschätzung des Prüfpersonals zur Veränderung des Gesundheitszustands des Studienteilnehmenden wird als nicht patientenrelevant bewertet. Grundsätzlich wird die Selbsteinschätzung der betroffenen Person zu ihrem Krankheitszustand im Vergleich zu einer Fremdeinschätzung für die Nutzenbewertung bevorzugt.

Validität

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

9-HPT – 9-Hole Peg Test

Der Endpunkt „9-HPT“ zur Erhebung der Feinmotorik der oberen Gliedmaßen wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

In der Studie MOXle Teil 2 wurde die Feinmotorik der oberen Gliedmaßen mit dem 9-Hole Peg Test (9-HPT) erhoben. Dabei handelt es sich um eine quantitative Messung der feinmotorischen Funktion der oberen Extremitäten (Arme und Hände).

Die zu testende Person sitzt an einem Tisch und soll 9 Stifte (Pegs) aus einem genormten Behälter herausnehmen und in die 9 Löcher eines Bretts stecken. Anschließend soll sie diese nacheinander wieder herausnehmen und in den Behälter zurücklegen. Es wird die Zeit gemessen, die für den Test benötigt wird. Es wird sowohl die dominante als auch die nicht-dominante Hand in 2 aufeinander folgenden Versuchen zweimal getestet. Längere Testzeiten spiegeln eine stärkere Beeinträchtigung der Funktion der oberen Extremitäten wider.

Verglichen werden soll die Einsteckgeschwindigkeit der beiden Gruppen. Es wird unterschieden zwischen der Geschwindigkeit je Peg (Pegs/Sekunden) und der Gesamtgeschwindigkeit (reziproker Wert = 1/Sekunden). Bei beiden Formeln wird der Wert je Hand berechnet, die Sekundenzahl der beiden Versuche wird gemittelt. Die primäre Analyse bezieht sich auf die Gesamtgeschwindigkeit der nicht-dominanten Hand (reziproker Wert). Neben den beschriebenen Analysen sollte eine explorative Responderanalyse durchgeführt werden mit der Verbesserung als Responsekriterium. Diese wurde im Studienbericht und Modul 4 nicht dargestellt.

Bewertung

Die Operationalisierung ist weitgehend nachvollziehbar. Der Erhebungsbogen liegt vor, allerdings fehlt das Studienmanual. Es liegen keine Angaben vor, welcher Wert bei einem Abbruch eingetragen wird. Es ist unklar, in welcher Reihenfolge der Test in Bezug auf die anderen Tests durchgeführt wurde.

Patientenrelevanz

Der Endpunkt 9-HPT wird im vorliegenden Anwendungsgebiet als patientenrelevant angesehen.

Validität

Es liegt eine Validierungsstudie mit 29 Personen mit FA vor, in der die konvergente Validität untersucht wurde. Diese zeigt eine hohe Korrelation mit der Krankheitsdauer, der Cooperative Ataxia Group ataxia scale und dem T25-FWT [16]. Darüber hinaus liegen mehrere Studien vor, in denen die psychometrischen Gütekriterien des Instruments bei anderen neurologischen Patientenpopulationen untersucht wurde (u. a. Morbus Parkinson, Multiple Sklerose). Es zeigen sich u. a. eine gute Reliabilität (Test-Retest- und Interrater-Reliabilität), Präzision, Konstruktvalidität und Änderungssensitivität [3,7,9,22]. Aufgrund der Validierung der 9-HPT in verschiedenen neurologischen Patientenpopulationen sowie dem vielfältigen Einsatz in klinischen Studien wird die Validität des Instruments als hinreichend übertragbar angesehen.

Im Studienbericht und Dossier wurde die Einsteckgeschwindigkeit dargestellt. Zur Beurteilung der Veränderung der Feinmotorik der oberen Gliedmaßen wird die Dauer (in Sekunden) als adäquater für eine inhaltlich sinnvolle und verständliche Interpretation angesehen.

T25-FWT – Timed 25 Foot Walk Test

Der Endpunkt „T25-FWT“ zur Erhebung der Beinfunktion wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

In der Studie MOXIe Teil 2 wurde die Beinfunktion mit dem Timed 25 Foot Walk Test (T25-FWT) erhoben. Beim T25-FWT wird die Zeit gemessen, die eine Person benötigt, um eine Strecke von 25 Fuß (7,6 m) zu bewältigen. Der Einsatz von Hilfsmitteln wie Gehhilfen und Rollatoren ist erlaubt. Wurde bei der ersten Erhebung ein Hilfsmittel verwendet, sollte dies auch bei allen weiteren Erhebungen genutzt werden. Längere Testzeiten spiegeln eine stärkere Beeinträchtigung der Gehfähigkeit der Person wider.

Laut SAP hatten die Personen 2 Versuche. Wenn beide Versuche erfolgreich waren, sollte der Mittelwert der benötigten Zeit gebildet werden. Andernfalls sollte nur die Zeit des abgeschlossenen Laufs für die Analysen verwendet werden. Verglichen werden sollte die Geschwindigkeit pro Fuß (Fuß/Sekunde) und die Gesamtgeschwindigkeit (reziproker Wert, 1/Sekunde). Die Gesamtgeschwindigkeit stellt gemäß SAP die Hauptanalyse dar. Außerdem

sollte eine Responderanalyse zum Anteil an Personen mit einer Verbesserung durchgeführt werden. Diese konnte im Studienbericht und Modul 4 nicht identifiziert werden.

Die Studienteilnehmenden konnten entscheiden, ob der Test aufgezeichnet werden soll oder nicht. Bei Zustimmung wurde der Test ein weiteres Mal ohne Zeitdruck und -messung durchgeführt. Für die Videoaufnahme gab es eine Anleitung.

Bewertung

Die beschriebene Operationalisierung des T25-FWT ist weitgehend nachvollziehbar. Es ist unklar, welcher Wert in den Erhebungsbogen eingetragen wurde, wenn eine Person den Test nicht geschafft hat. Das Studienmanual liegt nicht vor.

Patientenrelevanz

Der Endpunkt T25-FWT wird im vorliegenden Anwendungsgebiet als patientenrelevant angesehen.

Validität

Es liegt eine Validierungsstudie mit 29 Personen mit FA vor, in der die konvergente Validität untersucht wurde. Diese zeigt eine hohe Korrelation mit der Krankheitsdauer, der Cooperative Ataxia Group ataxia scale und dem 9-HPT [16]. Darüber hinaus liegen mehrere Übersichtsarbeiten zu den Gütekriterien des Instruments bei Personen mit Multiple Sklerose vor. Es zeigen sich u. a. eine gute Reliabilität und Konstruktvalidität. Die Variabilität der Gehgeschwindigkeit war im Allgemeinen bei Personen mit einer größeren Behinderung höher. Trainingseffekte haben sich nicht eingestellt [11,18]. Aufgrund der ähnlichen Symptomatik beim Laufen wird die Validität des Instruments als hinreichend übertragbar angesehen.

Im Studienbericht und Dossier wurde die Gehgeschwindigkeit dargestellt. In der vorliegenden Nutzenbewertung wird analog zum 9-HPT die Zeit in Sekunden als adäquater für eine inhaltlich sinnvolle und verständliche Interpretation angesehen.

ADL – Activities of Daily Living

Der Endpunkt „ADL“ zur Erhebung der Einschränkung in Alltagsaktivitäten wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

In der Studie MOXle Teil 2 dient der patientenberichtete Fragebogen „Activities of Daily Living (ADL)“ der Erhebung von Einschränkungen bei Alltagsaktivitäten.

Der Fragebogen enthält 9 krankheitsspezifische Items, die die Patientinnen und Patienten bezüglich ihrer Einschränkungen bei Tätigkeiten, Funktionen und Aktivitäten im täglichen Leben beurteilen sollen (Sprache, Schlucken, Essen schneiden und Handhabung von Utensilien, Ankleiden, persönliche Hygiene, Stürze, Gehen, Qualität der Sitzposition und Blasenfunktion). Ein niedriger Wert (0) bedeutet keine Einschränkung, ein hoher Wert (max. 4 Punkte), dass die Person nicht in der Lage ist, die Tätigkeit/Aktivität auszuführen. Der Gesamtscore addiert sich aus den Werten der einzelnen Items und reicht von 0 bis 36. Wenn die Bewertung von mindestens einem Item fehlt, wird der Gesamtscore nicht berechnet, sondern als fehlend betrachtet.

Die Erhebung erfolgte durch geschultes Prüfpersonal, wenn möglich sollte zu allen Erhebungszeitpunkten dieselbe Person den Test mit den Studienteilnehmenden durchführen.

Zur Auswertung soll die Veränderung gegenüber Baseline und der Anteil an Personen mit einer Verbesserung berechnet werden. Im Studienbericht wird definiert, dass eine klinisch relevante Verbesserung ab 0,4 Punkten pro Jahr vorliegt.

Bewertung

Die Operationalisierung ist weitestgehend nachvollziehbar. Laut pU liegt eine klinisch relevante Veränderung (MID) bei 0,4 Punkten pro Jahr. Die Herleitung der dargelegten MID kann anhand der Referenzpublikation [21] nicht nachvollzogen werden. Für die vorliegende Nutzenbewertung wird eine Relevanzschwelle von 15 % der Skalenspannweite als adäquat angesehen.

Patientenrelevanz

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

Validität

Detaillierte Informationen zur Fragebogenentwicklung der ADL-Skala liegen nicht vor und die Einbindung von Patientinnen und Patienten in den Entwicklungsprozess ist nicht dokumentiert. Die Inhaltsvalidität des Fragebogens erscheint vor diesem Hintergrund fraglich, jedoch weist die ADL-Skala hohe Korrelationen zu arztbewerteten Instrumenten und Funktionstests in der Indikation auf (FARS, 9-HPT, T25-FWT, Vision) [21,33]. Der ADL ist ein etabliertes Instrument und wird regelhaft in klinischen Studien eingesetzt [2,4,15,21]. Es liegen keine Informationen zur Reliabilität, Strukturvalidität und Änderungssensitivität vor. Basierend auf der Augenscheinvalidität wird das Instrument als geeignet angesehen, um Einschränkungen in Alltagsaktivitäten zu erfassen.

Häufigkeit von Stürzen

Der Endpunkt „Häufigkeit von Stürzen“ wird in der Nutzenbewertung aufgrund unklarer Validität nicht berücksichtigt.

Operationalisierung

Beschreibung

In der Studie MOXle Teil 2 wurde ein Sturz definiert als ein Ereignis, bei dem der oder die Betroffene unbeabsichtigt auf dem Boden oder auf einer anderen tieferen Ebene zum Liegen aufkommt.

Die Probanden erhielten während des Screenings ein sogenanntes „Sturz-Tagebuch“ und sollten alle Stürze, die zwischen dem Screening und der Visite in Woche 48 auftraten, eintragen. Zu jedem Sturzeintrag gehörte u. a. das Datum und die Uhrzeit eines jeden Sturzes, der Ort des Sturzes, die dem Sturz vorangegangene Aktivität, die wahrgenommene Ursache des Sturzes sowie eventuelle Verletzungen nach dem Sturz.

Die Gesamtzahl der Stürze vor und nach der Verabreichung des Studienmedikaments wurde angegeben sowie die Anzahl der Stürze pro 12 Wochen-Intervall (d. h. Tag 1 bis Woche 12, Woche 12 bis 24, Woche 24 bis 36 und Woche 36 bis 48) zusammengefasst und nach Anzahl der Patientinnen und Patienten pro Gruppe normiert. Anschließend wurde die Anzahl der Stürze deskriptiv zusammengefasst.

Bewertung

Die Operationalisierung ist weitgehend nachvollziehbar. Es wurde nicht beschrieben, ob das Sturz-Tagebuch in digitaler Form oder in Papierform war. Aufgrund der möglichen feinmotorischen und visuellen Einschränkungen bei FA ist unklar, ob alle Patientinnen und Patienten in der Lage waren, das Tagebuch selbstständig zu führen. Eine mögliche Unterstützung durch betreuende Angehörige wurde im Protokoll nicht erwähnt. Auch wurden ein vorheriges Training oder eine Anleitung zum standardisierten Vorgehen nicht beschrieben. Darüber hinaus fehlt die Information, in welchem zeitlichen Intervall das Sturztagebuch (z. B.

direkt nach dem Sturz, einmal täglich oder in einem anderen zeitlichen Rhythmus) ausgefüllt werden sollte. Daher ist unklar, ob die Daten vollständig erhoben wurden.

Patientenrelevanz

Stürze werden, insbesondere aufgrund der schwerwiegenden Verletzungen, die in Folge eines unkontrollierten Sturzes eintreten können, per se als patientenrelevant eingestuft. Das Risiko für Stürze ist abhängig vom Schweregrad der Friedreich-Ataxie. Patienten und Patientinnen mit einer stärker ausgeprägten Gangstörung stürzen gegebenenfalls häufiger.

Validität

Aufgrund der bei der Operationalisierung genannten Kritikpunkte und Unsicherheiten kann die Validität der Erhebung der Stürze nicht bewertet werden.

2.3.3 Lebensqualität

SF-36

Der Endpunkt „SF-36“ wird in der Nutzenbewertung berücksichtigt.

Operationalisierung

Beschreibung

Der Fragebogen „Short-Form-36 Health Survey“ (SF-36) wurde in der Studie MOXle Teil 2 zur Erfassung der selbstberichteten gesundheitsbezogenen Lebensqualität eingesetzt.

Der SF-36 beinhaltet 36 Items, die 8 Domänen zugeordnet werden: „Körperliche Funktionsfähigkeit“, „Körperliche Rollenfunktion“, „Körperlicher Schmerz“, „Allgemeine Gesundheitswahrnehmung“, „Vitalität“, „Soziale Funktionsfähigkeit“, „Emotionale Rollenfunktion“ und „Psychisches Wohlbefinden“. Zusätzlich werden die 8 Domänen zu einem „Körperlichen Summenscore“ (Physical Component Summary, PCS) und einem „Psychischen Summenscore“ (Mental Component Summary, MCS) zusammengefasst. Ausgewertet werden sowohl die Domänenwerte als auch die Summenscores. Für die Domänen- und Summenscores bedeuten höhere Werte eine bessere gesundheitsbezogene Lebensqualität. Bezugszeitraum ist die vergangene Woche [35].

Eine Verschlechterung wurde als Veränderung von Baseline zu Woche 48 um $\leq -9,4$ Punkte in der PCS und $\leq -9,6$ Punkte in der MCS definiert. Dies entspricht 15 % der Skalenspannweite.

Bewertung

Die Operationalisierung ist nachvollziehbar. Die Responseschwelle für eine Verschlechterung von $\leq -9,4$ bzw. $\leq -9,6$ Punkten ist adäquat.

Patientenrelevanz

Die Lebensqualität mittels SF-36 wird in der vorliegenden Operationalisierung als patientenrelevanter Endpunkt entsprechend § 2 Satz 3 AM-NutzenV angesehen.

Validität

Der Endpunkt SF-36 wird als generisches und valides Erhebungsinstrument zur Erfassung der gesundheitsbezogenen Lebensqualität bewertet.

2.3.4 Sicherheit

Unerwünschte Ereignisse

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

Operationalisierung

Beschreibung

Unerwünschte Ereignisse (UE) sind in der Studie MOXle Teil 2 definiert als jegliche unerwünschte medizinische Ereignisse, unabhängig davon, ob sie als arzneimittelbedingt angesehen werden oder nicht. Ein UE kann jedes ungünstige und unbeabsichtigte Anzeichen (z. B. anormaler Laborbefund), Symptom oder jede Krankheit sein, die zeitlich mit der Anwendung eines Arzneimittels zusammenhängt, ohne dass ein Urteil über die Kausalität möglich ist. Diese Definition schließt jedes neu aufgetretene Ereignis oder jeden früheren Zustand ein, der seit der Verabreichung des Studienmedikaments an Schwere oder Häufigkeit zugenommen hat.

UE, die auf Begleiterkrankungen, Reaktionen auf Begleiterkrankungen, Reaktionen auf Begleitmedikationen oder das Fortschreiten von Krankheitszuständen zurückzuführen sind, müssen ebenfalls gemeldet werden. Vorerkrankungen (die vor Beginn des Erfassungszeitraums der UE vorliegen) gelten als Begleiterkrankungen und sollten nicht als UE erfasst werden. Tritt bei der Patientin / dem Patienten jedoch eine Verschlechterung oder Komplikation einer solchen Begleiterkrankung auf, sollte die Verschlechterung oder Komplikation (mit Ausnahme des Fortschreitens der Erkrankung) als UE erfasst werden.

Jedes UE sollte als eine einzige Diagnose erfasst werden. Begleitende Anzeichen (einschließlich abnormaler Labortestwerte oder EKG-Befunde) oder Symptome sollten nicht als zusätzliche UE erfasst werden. Wenn eine Diagnose nicht bekannt ist, sollten die Anzeichen oder Symptome als UE erfasst werden. Veränderungen der Labortestwerte oder EKG-Parameter gelten nur dann als UE, wenn sie als klinisch bedeutsam eingestuft werden. Wenn abnormale Labortestwerte oder EKG-Befunde die Folge einer Pathologie sind, für die es eine Gesamtdiagnose gibt (z. B. erhöhte Kreatininwerte bei Nierenversagen), sollte nur die Diagnose als UE gemeldet werden.

Geplante elektive Eingriffe (Operationen oder Therapien), die zur Behandlung von Erkrankungen durchgeführt werden, die bereits vor der Teilnahme der Person an der Studie bestanden (z. B. elektive Parodontaloperationen, elektive Leistenbruchoperationen), sollten nicht als UE erfasst werden, sondern nur in den Originalunterlagen des Studienteilnehmenden dokumentiert werden. Wenn ein zuvor geplanter Eingriff vorzeitig durchgeführt wird (z. B. als Notfall), weil sich der zuvor bestehende Zustand verschlechtert, sollte die Verschlechterung des Zustands als UE erfasst werden.

In der Studie MOXle Teil 2 sollten UE ab der Einnahme der Studienmedikation bis 30 Tage nach der letzten Gabe der Studienmedikation erhoben werden.

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

Das Auftreten von UE wird gemäß Medical Dictionary for Regulatory Activities (MedDRA, 14.1) nach Systemorganklasse und Preferred Term erfasst. In der Studie wird der Schweregrad des UE wie folgt eingeteilt:

- Mild: Symptome, die die üblichen sozialen und funktionellen Aktivitäten nicht oder nur minimal beeinträchtigen
- Moderat: Symptome, die eine mehr als minimale Beeinträchtigung der üblichen sozialen und funktionellen Aktivitäten verursachen
- Schwer: Symptome, die zur Unfähigkeit führen, übliche soziale und funktionelle Aktivitäten durchzuführen.

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

- Tod
- Lebensbedrohlich
- Bedurfte einer Hospitalisierung oder der Verlängerung einer bestehenden Hospitalisierung
- Führt zu anhaltender oder signifikanter Behinderung oder erheblichen Beeinträchtigungen, normale Lebensfunktionen auszuüben
- Kongenitale Anomalie / ein Geburtsfehler
- Wichtiges medizinisches Ereignis (Ereignisse, die keines der oben definierten Kriterien erfüllen. Sie können jedoch als schwerwiegend angesehen werden, wenn sie nach angemessener medizinischer Beurteilung die Person gefährden können und einen medizinischen oder chirurgischen Eingriff erfordern, um eines der anderen in der SUE-Definition aufgeführten Ereignisse zu verhindern.)

Bewertung

Die Operationalisierung ist nachvollziehbar. Die studienindividuelle Schweregradeinteilung der UE wird als sachgerecht beurteilt.

Patientenrelevanz und Validität

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

2.3.5 Erhebungszeitpunkte

Eine Übersicht der Erhebungszeitpunkte der in der Nutzenbewertung berücksichtigten Endpunkte findet sich in Tabelle 5.

Tabelle 5: Erhebungszeitpunkte der berücksichtigten Endpunkte in der Studie MOXie Teil 2

Studienvisite (Woche) Endpunkt	Visite 1 S	Visite 2 Tag 1	Visite 7 W12	Visite 9 W24	Visite 11 W36	Visite 13 W48 (Behandlungs- ende)	Visite 14 W52 ¹⁾ (Studienende/4- wöchige Nach- beobachtung)
Todesfälle		kontinuierlich					
PGI-C ²⁾			x	x	x	x	
9-HPT	x	x		x		x	
T25-FWT	x	x ³⁾		x ³⁾		x ³⁾	
ADL	x			x	x	x	
SF-36		x		x		x	
Unerwünschte Ereignisse		kontinuierlich					

¹⁾ Untersuchungen sollten auch im Falle eines vorzeitigen Studienabbruchs durchgeführt werden.

²⁾ Patientinnen und Patienten füllten den PGI-C im Anschluss an die neurologische FARS- oder CGI-C-Erhebung aus.

³⁾ Bei Studienteilnehmenden, die einer Videoaufzeichnung des T25-FWT zustimmten, wurde der Test ein weiteres Mal ohne Zeitmessung durchgeführt.

Abkürzungen: 9-HPT: 9-Hole Peg Test; ADL: Activities of Daily Living; CGI-C: Clinical Global Impression of Change; FARS: Friedreich Ataxia Rating Scale; PGI-C: Patient Global Impression of Change; S: Screening; SF-36: Short-Form-36 Health Survey; T25-FWT: Timed 25-Foot Walk Test; W: Woche.

2.4 Statistische Methoden

Der ursprüngliche SAP (Version 1.0) wurde am 16.05.2019 veröffentlicht und basiert auf dem Studienprotokoll Version 9.0 (17.08.2017). Die Veröffentlichung des finalen SAP (Version 2.0) erfolgte am 18.09.2019 und basiert auch auf dem Studienprotokoll Version 9.0.

Analysepopulationen

- ITT-Population: umfasst alle randomisierten Personen, kategorisiert nach ihrer randomisierten Behandlungszuweisung (unabhängig davon, ob sie das Studienmedikament erhalten haben oder nicht). In den originalen Studiendokumenten wird diese Population als „All Randomized Population (ARP)“ bezeichnet und war für deskriptive Analysen der Wirksamkeit vorgesehen.
- Full Analysis Set (FAS): Alle Personen ohne Pes cavus, für die mindestens ein Post-Baseline-Wert vorliegt, kategorisiert nach der jeweiligen randomisierten Behandlungszuweisung (unabhängig davon, ob sie das Studienmedikament erhalten haben oder nicht). Die FAS-Population war für die primäre Analyse der Wirksamkeit vorgesehen.
- Sicherheitspopulation: Alle Personen, die mindestens eine Dosis der randomisierten Studienmedikation erhalten haben. Die Sicherheitspopulation wird verwendet für die Analyse der Sicherheitsendpunkte.

Datenschnitte

Für die Nutzenbewertung wird der finale Datenschnitt verwendet. Für die Wirksamkeitsendpunkte werden die Auswertungen (ITT-Population) zu Woche 48 und für die Sicherheitsendpunkte die Auswertungen (Sicherheitspopulation) zu Woche 52 herangezogen.

Präspezifizierte Subgruppenanalysen

Gemäß SAP waren für alle Endpunkte folgende Subgruppen präspezifiziert:

- Alter (< 18 vs. ≥ 18 Jahre)
- Geschlecht (weiblich vs. männlich)
- Geografische Region (USA vs. andere)
- Ethnizität (nicht-hispanisch/-lateinamerikanisch vs. hispanisch/lateinamerikanisch)
- Abstammung (weiß vs. andere)
- GAA1-Repeat-Länge ≥ 675 (ja vs. nein)

Es liegen keine Interaktionstests für Subgruppen vor. Einzig für die GAA1-Repeat-Länge ≥ 675 wurden Subgruppenauswertungen für den PGI-C (ITT-Population) im Modul 4 präsentiert. Zudem gibt es keine Angaben zu fehlenden Werten in den jeweiligen Subgruppen.

Präspezifizierung geplanter und durchgeführter Analysen

Zur Bestimmung der LS-Mean Differenz zwischen den beiden Behandlungsgruppen für den PGI-C (nach 48 Wochen) war eine Kovarianzanalyse (ANCOVA) präspezifiziert. Post hoc wurde eine Responderanalyse für eine Verbesserung (< 4 Punkte, jegliche Verbesserung) bzw. Verschlechterung (> 4 Punkte; jegliche Verschlechterung) im PGI-C zu Woche 48 vorgelegt. Fehlende Werte wurden nicht imputiert. Für die Nutzenbewertung werden die Ergebnisse der Responderanalyse dargestellt.

Zur Bestimmung der LS-Mean Differenz zwischen den beiden Behandlungsgruppen für die Endpunkte ADL, 9-HPT, T25-FWT und SF-36 (nach 48 Wochen) waren Analysen mittels Mixed Model for Repeated Measurement (MMRM) geplant. Post hoc wurde eine Responderanalyse für

eine Verschlechterung im SF-36 mit einer Relevanzschwelle von $\leq -9,4$ bzw. $\leq -9,6$ Punkten vorgelegt, die in der Nutzenbewertung dargestellt wird.

Für die Sicherheitsendpunkte wurden post hoc relative Risiken (RR) mit 95%-Konfidenzintervall (95%-KI) berechnet.

Fehlende Werte und Daten-Imputation

In der Studie MOXle Teil 2 wurden keine Imputationen für die in der Nutzenbewertung herangezogenen Endpunkte durchgeführt.

Einschätzung der statistischen Auswertungen

Die durchgeführten statistischen Analysen werden als adäquat angesehen. Es fehlen jedoch Angaben zum verwendeten statistischen Modell bzw. Testverfahren und zu Adjustierungsvariablen in der Auswertung der unerwünschten Ereignisse.

2.5 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 6: Verzerrungspotential der Studie MOXle Teil 2

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung Patientin/Patient	Verblindung Behandlungsperson	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential auf Studienebene
MOXle Teil 2	Ja	Ja	Unklar ¹⁾	Unklar ¹⁾	Nein	Ja ²⁾³⁾	Hoch

¹⁾ Unbeabsichtigte Entblindung durch das häufige Auftreten charakteristischer UE im Omaveloxolon-Arm wie beispielsweise „Übelkeit“ (33,3 %), „Abdominalschmerz“ (21,6 %), „ALT erhöht“ (37,3 %) oder „AST erhöht“ (21,6 %) möglich.

²⁾ Es bestehen Unterschiede in den Baselinecharakteristika zwischen den Behandlungsgruppen: 60,8 % der Studienteilnehmenden im Omaveloxolon-Arm waren weiblich, im Placebo-Arm waren lediglich 32,7 % weiblich. Der Anteil an Personen mit einer GAA1-Repeat-Länge ≥ 675 war höher im Omaveloxolon-Arm (51,0 %) als im Placebo-Arm (40,4 %). Studienteilnehmende im Omaveloxolon-Arm wiesen häufiger eine Kardiomyopathie (49,0 %) sowie operative Behandlung der Skoliose (31,4 %) als im Placebo-Arm (28,8 % bzw. 19,2 %) auf.

³⁾ Unterschied von ca. 10 % im vorzeitigen Abbruch der Studie zwischen den Behandlungsarmen (Omaveloxolon: 11,8 % und Placebo: 1,9 %).

Abkürzungen: ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; GAA: Guanin-Adenin-Adenin; UE: Unerwünschtes Ereignis.

Das Verzerrungspotential auf Studienebene wird als hoch eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 7 dargestellt.

Tabelle 7: Verzerrungspotential der in der Nutzenbewertung berücksichtigten Endpunkte der Studie MOXle Teil 2

Endpunkt	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige, das Verzerrungspotential beeinflussende Punkte	Verzerrungspotential
Todesfälle	Ja	Nein ¹⁾²⁾	Nein	Nein	Unklar ³⁾
PGI-C	Unklar ⁴⁾	Nein ⁵⁾	Nein	Unklar ⁶⁾	Hoch
ADL	Unklar ⁴⁾	Nein ⁵⁾	Nein	Unklar ⁷⁾	Hoch
SF-36	Unklar ⁴⁾	Nein ⁵⁾	Nein	Unklar ⁷⁾	Hoch
Unerwünschte Ereignisse	Unklar ⁴⁾	Nein ¹⁾²⁾	Nein	Unklar ⁷⁾⁸⁾	Hoch

¹⁾ Auswertungen in Sicherheitspopulation.

²⁾ 6 Personen (11,8 %) im Omaveloxolon-Arm und 1 Person (1,9 %) im Placebo-Arm brachen die Studie vorzeitig ab. Zudem liegen keine Angaben zur Beobachtungszeit vor.

³⁾ Da in beiden Behandlungsarmen die gleiche Häufigkeit an Todesfällen (N = 0) berichtet wurde, wird die Ergebnissicherheit als nicht eingeschränkt angesehen.

⁴⁾ Es kann nicht ausgeschlossen werden, dass das häufige Auftreten charakteristischer UE im Omaveloxolon-Arm wie beispielsweise „Übelkeit“ (33,3 %), „Abdominalschmerz“ (21,6 %), „ALT erhöht“ (37,3 %) oder „AST erhöht“ (21,6 %) zu einer Entblindung der betroffenen Personen und Behandler geführt hat.

⁵⁾ Zu Woche 48 liegen Daten für 44 Personen (86,3 %) im Omaveloxolon-Arm und für 51 Personen (98,1 %) im Placebo-Arm vor. Fehlende Daten wurden nicht imputiert. Aufgrund des Unterschiedes zwischen den Behandlungsarmen von 11,8 % in den Rücklaufquoten zu Woche 48 wird das Verzerrungspotential als hoch eingestuft.

⁶⁾ Trotz Unterschieden > 10 % zwischen den Behandlungsarmen wurden die Variablen Geschlecht und Kardiomyopathien nicht im statistischen Modell bzw. in Sensitivitätsanalysen berücksichtigt.

⁷⁾ Relevante Confounder (siehe Kapitel 2.6 indirekter Vergleich) wie z. B. das Geschlecht und GAA1-Repeat-Länge ≥ 675 sowie Kardiomyopathien wurden trotz Unterschieden > 10 % zwischen den Behandlungsarmen nicht im statistischen Modell bzw. in Sensitivitätsanalysen berücksichtigt.

⁸⁾ Es konnten keine Angaben zur Berechnung des Effektschätzers und der p-Werte identifiziert werden und es bleibt unklar, ob das Stratifizierungsmerkmal Pes cavus (Pes cavus vs. kein Pes cavus) im statistischen Modell berücksichtigt wurde.

Abkürzungen: ADL: Activities of Daily Living; ALT: Alanin-Aminotransferase; AST: Aspartat-Aminotransferase; GAA: Guanin-Adenin-Adenin; ITT: Intention to treat; PGI-C: Patient Global Impression of Change; SF-36: Short-Form-36 Health Survey; UE: Unerwünschtes Ereignis.

2.6 Indirekter Vergleich

Der pU legt mit dem Dossier als supportive Evidenz für den primären Endpunkt „modifizierte Friedreich Ataxia Rating Scale (mFARS)“ einen indirekten Vergleich zur Langzeitwirksamkeit von Omaveloxolon gegenüber „Best Supportive Care“ (BSC) in Form einer Propensity-Score-Matching-Analyse über einen Zeitraum von 3 Jahren vor [13,17]. Hierbei wurde der Verlauf der Krankheitsschwere gemessen anhand der mFARS aus der Open-Label-Extensionsphase (OLE) der MOXle-Studie in einem indirekten Vergleich ohne Brückenkomparator einer natürlichen Verlaufskohorte aus der „Friedreich Ataxia Clinical Outcome Measures Study (FA-COMS)“ als externe Kontrollkohorte gegenübergestellt. Der pU reicht für diese Post-hoc-Analyse einen SAP und Studienbericht für den indirekten Vergleich ein [26,27]. In den eingereichten Unterlagen konnte kein Studienprotokoll für das Register FA-COMS identifiziert werden.

Um in die OLE der MOXle-Studie eingeschlossen zu werden, mussten die Patientinnen und Patienten eine 12-wöchige Behandlung in Teil 1 oder eine 48-wöchige Behandlung mit Omaveloxolon in Teil 2 abgeschlossen haben, durften keine Protokollabweichungen aufweisen und mussten die Einschluss- und Ausschlusskriterien der OLE erfüllen. Alle Studienteilnehmenden der OLE erhielten einmal täglich 150 mg Omaveloxolon. Für die Studie MOXle OLE wurde der Datenschnitt vom 24. März 2022 der 144-wöchigen OLE für den indirekten Vergleich berücksichtigt und umfasste n = 149 Studienteilnehmende. Für FA-COMS wurden Daten vom 24. März 2021 und n = 810 Personen genutzt.

Als externe Kontrollkohorte wurden retrospektiv patientenindividuelle Daten aus dem Register FA-COMS herangezogen. Den Studienunterlagen ist zu entnehmen, dass in FA-COMS Männer oder Frauen im Alter von 16 bis 40 Jahren und einer genetisch bestätigten Friedreich-Ataxie (FA) eingeschlossen wurden. Ausschlusskriterien waren nicht definiert. Die FA-Patientinnen und Patienten wurden in 14 Studienzentren (USA: N = 9, Australien, Neuseeland & Indien: N = 1; Kanada: N = 2) seit 2003 rekrutiert. In der FA-COMS erfolgte die Erhebung des mFARS einmal im Jahr, in der MOXle OLE alle 24 Wochen (ca. 2 x Jahr). Darüber hinaus wird für den indirekten Vergleich (zwischen FA-COMS und Studie MOXle OLE) anstatt des validierten mFARS mit 93 Punkten eine Version mit 99 Punkten verwendet. Zu einer Version des mFARS mit 99 Punkten finden sich keine Angaben in der Literatur, so dass der Endpunkt als nicht adäquat angesehen wird.

Um in die jeweilige Studienpopulation für das Propensity Score Matching aufgenommen werden zu können, mussten die Patientinnen und Patienten in der MOXle OLE und FA-COMS einen mFARS-Baselinewert, mindestens eine Post-Baseline-mFARS innerhalb von 3 Jahren nach Baseline und Werte für alle Propensity-Score-Modell-Kovariaten (Geschlecht, Baselinewert mFARS-Score, Alter, Alter bei Beginn der FA sowie Baselinewert für die Gangart) aufweisen. Die Auswahl geeigneter Personen auf Grundlage der Verfügbarkeit von Daten erhöht das Risiko für einen Selektionsbias. Für den indirekten Vergleich wurden jeweils n = 136 Personen eingeschlossen. Für eine Sensitivitätsanalyse (Subpopulation) mussten die Patientinnen und Patienten der externen Kontrollgruppe (FA-COMS) folgende zusätzlichen Anforderungen erfüllen: mFARS-Baselinewert von 8 bis 74 und Alter zwischen 16 und 40 Jahren (n = 278); analog der Einschlusskriterien in der MOXle-Studie.

Der indirekte Vergleich wird aufgrund verschiedener Limitationen nicht in der Nutzenbewertung herangezogen. Ein wesentlicher Kritikpunkt ist, dass keine systematische Recherche sowie Expertenbefragung für die Auswahl der Confounder erfolgte. Es wurden lediglich Faktoren, die sich als prognostisch erwiesen haben und in beiden Studien (FA-COMS und MOXle OLE) vorhanden waren, als Confounder zur Bestimmung der Propensity-Scores verwendet. Die Auswahl von Confoundern anhand der Verfügbarkeit im Datensatz wird als inadäquat bewertet, da ein als relevant identifizierter Confounder seinen Stellenwert nicht dadurch verliert, dass zu ihm keine Informationen im Datensatz vorliegen. Für die Aufnahme in die Vergleichspopulation mussten in der FA-COMS Daten für die verwendeten Confounder vorliegen, aber nicht für wesentliche Ein- und Ausschlusskriterien der klinischen Studie (u. a. kardiovaskuläre Komorbiditäten). Dies schränkt eine mögliche Validität und Verallgemeinerbarkeit eines adjustierten Propensity-Score-Modells ein.

Insgesamt gingen 5 Confounder (Geschlecht, Baselinewert mFARS-Score, Alter, Alter bei Beginn der FA sowie Baselinewert für die Gangart) in das Modell zur Herleitung der Propensity Scores durch eine logistische Regression ein. Die beiden Variablen „Guanin-Adenin-Adenin (GAA1)-Repeat-Länge“ und „Pes-Cavus-Status“ waren laut pU nicht für alle Patientinnen und Patienten verfügbar und wurden daher nicht berücksichtigt in der Propensity-Score-Matching-Analyse. Die GAA1-Repeat-Länge lag bei 92 % der FA-COMS- und 87,9 % der MOXle-OLE-

Teilnehmenden vor. Der Pes-cavus-Status hingegen lag nur bei 53,3 % der FA-COMS-Teilnehmenden und bei allen Teilnehmenden der MOXle OLE (Stratifizierungsfaktor der Randomisierung für MOXle Teil 2) vor. Da sich statistisch signifikante Unterschiede ($p \leq 0,0001$) in der GAA1-Repeat-Länge zwischen den Studienteilnehmenden der FA-COMS und MOXle OLE zeigen, kann von keiner Strukturgleichheit der beiden Populationen ausgegangen werden. Beide Variablen werden in der MOXle Teil 2 (RCT) mittels Subgruppenanalyse (für GAA1-Repeat-Länge) sowie der Analysen im FAS (Personen ohne Pes cavus) vertiefter ausgewertet. Des Weiteren zeigen sich auch in der GAA2-Repeat-Länge statistisch signifikante Unterschiede ($p \leq 0,0001$) zwischen den Studienteilnehmenden der FA-COMS und MOXle OLE. Für diese 3 Variablen liegen keine Angaben zur SMD vor. Die Berücksichtigung aller relevanten Confounder und Effektmodifikatoren als Adjustierungsvariablen ist bei der Durchführung eines indirekten Vergleichs ohne Brückenkomparator notwendig, um die notwendige Strukturgleichheit zu erreichen. Das beschriebene Vorgehen zur Confounder-Identifikation stellt jedoch nicht sicher, dass relevante bzw. nicht gemessene Faktoren als Confounder berücksichtigt wurden. Zudem werden relevante Unterschiede bekannter Confounder bei der Adjustierung nicht berücksichtigt.

Darüber hinaus schließen in MOXle Teil 2 von $N = 51$ Personen im Interventionsarm nur $n = 44$ den 48-wöchigen RCT ab und $n = 43$ gehen in die OLE über (84 %). Im Placeboarm hingegen schließen von $N = 52$ Personen $n = 50$ den RCT ab und $n = 49$ gehen in die OLE über (94 %). Dies führt zu einem Selektionsbias (Survivorship-Bias), so dass die Population in der OLE nicht mehr der ITT-Population entspricht. Zudem beginnt die Datenerhebung in FA-COMS im Jahr 2003, der erste Patient / die erste Patientin wurde in die MOXle OLE am 30. Oktober 2018 eingeschlossen, so dass sich ein zeitlicher Unterschied von 15 Jahren ergeben kann. Durch den langen Zeitraum ist unsicher, ob der Therapiealgorithmus (z. B. Diagnostik, Physiotherapie) in diesem heterogenen Krankheitsbild über den Zeitraum hinweg vergleichbar mit der Einschätzung zum Zeitpunkt der Studiendurchführung geblieben ist. Die fehlende zeitliche Parallelität der klinischen Studie und des Registers kann Auswirkungen auf die Positivität haben.

In den eingereichten Unterlagen werden die Ergebnisse des indirekten Vergleichs, inklusive Propensity-Score-Matching zum mFARS beschrieben. Es sind jedoch nur unzureichend Informationen zur Güte des indirekten Vergleichs, wie Gegenüberstellung der Baselinecharakteristika der Grundpopulationen (810 Personen der FA-COMS und 149 Personen der MOXle OLE) oder auch zur Überprüfung der Robustheit des Verfahrens verfügbar (u. a. Sensitivitätsanalysen).

2.6.1 Zusammenfassende Bewertung der Methodik des indirekten Vergleichs

Der indirekte Vergleich ohne Brückenkomparator wird nicht für die Nutzenbewertung berücksichtigt. Ein wesentlicher Kritikpunkt bezieht sich auf die Confounderidentifikation einschließlich der fehlenden systematischen Recherche nach relevanten Faktoren sowie auf die Frage der Strukturgleichheit zwischen den vergleichenden Populationen durch potentiell nicht adäquat abgebildete Ein- und Ausschlusskriterien. Zudem besteht ein Risiko für Selektionsbias aufgrund der Studienabbrecher insbesondere im Omaveloxolon-Arm in Teil 2 der MOXle-Studie, so dass die Population in der OLE nicht mehr der ITT-Population entspricht. Eine weitere Limitation ist die nicht adäquate Operationalisierung der mFARS und die fehlende zeitliche Parallelität von FA-COMS und MOXle OLE.

3 Ergebnisse der eingeschlossenen Studie

3.1 Studiencharakteristika und Studienmedikation

Tabelle 8: Allgemeine Angaben; Studie MOXle Teil 2

Studie MOXle Teil 2	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)
ITT-Population ¹⁾	51 (100)	52 (100)
FAS ¹⁾	40 (78,4)	42 (80,8)
Sicherheitspopulation ¹⁾	51 (100)	52 (100)
Behandlung bis Woche 48 abgeschlossen	44 (86,3)	50 (96,2)
Abbruch der Einnahme der Studienmedikation Aufgrund von:	7 (13,7)	2 (3,8)
UE	4 (7,8)	2 (3,8)
Entzug der Einverständniserklärung	3 (5,9)	0 (0)
Studien-Follow-up bis Woche 52 abgeschlossen	45 (88,2)	51 (98,1)
Abbruch der Studie Aufgrund von:	6 (11,8)	1 (1,9)
Administrativen Gründe	2 (3,9)	0 (0)
Entzug der Einverständniserklärung	4 (7,8)	1 (1,9)
Mediane Behandlungsdauer in Tagen (min; max)	336,0 (17; 355)	337,0 (57; 343)
Mediane Beobachtungsdauer in Tagen (min; max)	k. A.	k. A.

¹⁾ Definition, siehe Kapitel 2.4 Statistische Methoden

Abkürzungen: FAS: Full Analysis Set; ITT: Intention to treat; k. A.: keine Angabe; UE: Unerwünschtes Ereignis.

Tabelle 9: Charakterisierung der Studienpopulation der Studie MOXle Teil 2; ITT-Population

Studie MOXle Teil 2	Omaveloxolon N = 51	Placebo N = 52
<i>Alter (Jahre)</i>		
MW (SD)	23,4 (6,1)	24,1 (7,9)
Median (min; max)	22,0 (16; 39)	21,0 (16; 40)
<i>Altersgruppe (Jahre), n (%)</i>		
< 18	9 (17,6)	15 (28,8)
≥ 18	42 (82,4)	37 (71,2)
<i>Geschlecht, n (%)</i>		
männlich	20 (39,2)	35 (67,3)
weiblich	31 (60,8)	17 (32,7)
<i>Abstammung (genetisch), n (%)</i>		
kaukasisch/weiß	50 (98,0)	50 (96,2)
andere	1 (2,0)	2 (3,8)
<i>Region, n (%)</i>	k. A.	k. A.
<i>Pes Cavus¹⁾, n (%)</i>		
ja	10 (19,6)	10 (19,2)
nein	41 (80,4)	42 (80,8)
<i>Alter bei Beginn der FA (Jahre)</i>		
MW (SD)	14,8 (5,7)	15,3 (5,3)
Median (min; max)	13,0 (5; 36)	15,5 (5; 30)

Studie MOXle Teil 2	Omaveloxolon N = 51	Placebo N = 52
<i>BMI</i>		
MW (SD)	23,2 (5,1)	22,8 (4,8)
Median (min; max)	22,7 (16,1; 43,0)	21,4 (15,6; 40,5)
<i>Krankheitsdauer (Jahre)</i>		
MW (SD)	4,7 (3,8)	4,4 (4,4)
Median (min; max)	4,0 (0; 16)	3,0 (0; 17)
<i>Durchschnittliche Anzahl an Stunden pro Woche, an denen ein Trainingsprogramm (Exercise Regimen) durchgeführt wurde</i>		
MW (SD)	5,8 (4,5)	5,1 (4,3)
Median (min; max)	5,0 (0; 28)	4,0 (0; 17)
<i>GAA1-Repeat-Länge²⁾</i>		
MW (SD)	n = 41 736,8 (206,8)	n = 43 676,2 (267,9)
Median (min; max)	700,0 (230; 1160)	670,0 (170; 1270)
<i>GAA1-Repeat-Länge ≥ 675²⁾, n (%)</i>		
nein	n = 41 15 (29,4)	n = 43 22 (42,3)
ja	26 (51,0)	21 (40,4)
<i>GAA2-Repeat-Länge²⁾</i>		
MW (SD)	n = 40 781,5 (310,8)	n = 42 762,4 (316,5)
Median (min; max)	750,0 (6; 1500)	752,5 (170; 1400)
<i>Gehfähigkeit, n (%)</i>		
gefähig	46 (90,2)	49 (94,2)
nicht gefähig	5 (9,8)	3 (5,8)
<i>Mobilitätshilfen, n (%)</i>		
Keine	25 (49,0)	29 (55,8)
Rollstuhl	11 (21,6)	8 (15,4)
Gehhilfe	10 (19,6)	8 (15,4)
Gehstock	2 (3,9)	3 (5,8)
Hilfe durch eine andere Person	0 (0)	1 (1,9)
Hand der Betreuungsperson halten	1 (2,0)	0 (0)
<i>In Situationen, die Gleichgewicht erfordern:</i>		
Unterstützung durch eine andere Person oder einen Handtrolley	1 (2,0)	0 (0)
Mobilitätsroller und Rollstuhl	0 (0)	1 (1,9)
Rolling bag oder Rollator	0 (0)	1 (1,9)
Benutzt Scooter, gelegentlich auch Gehhilfe und Rollstuhl	1 (2,0)	0 (0)
Wanderstock	0 (0)	1 (1,9)
<i>Krankheitsgeschichte, n (%)</i>		
Kardiomyopathie	25 (49,0)	15 (28,8)
Areflexie	47 (92,2)	51 (98,1)
Operativer Eingriff am Fuß	4 (7,8)	2 (3,8)
Sensorische Neuropathie	26 (51,0)	26 (50,0)
Schwierigkeiten beim Schlucken	12 (23,5)	17 (32,7)
Skoliose	39 (76,5)	37 (71,2)
Operative Behandlung der Skoliose	16 (31,4)	10 (19,2)

¹⁾ Stratifikationsfaktor der Randomisierung.

²⁾ Für die Teilnahme an der Studie musste eine genetisch bestätigte Diagnose der FA vorliegen. Das Vorliegen von Informationen über die Länge der GAA-Repeats ist auf die jeweilige Methode und Dokumentation in der Krankenakte beschränkt.

Abkürzungen: FA: Friedreich-Ataxie; GAA: Guanin-Adenin-Adenin; ITT: Intention to treat; k. A.: keine Angabe; MW: Mittelwert; SD: Standardabweichung.

Protokollverletzungen

Anhand der in der Studie MOXle Teil 2 dokumentierten Protokollverletzungen ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

Begleitmedikation

Anhand der in der Studie MOXle Teil 2 dokumentierten Begleitmedikation ergeben sich keine verzerrenden Aspekte für die Bewertung des Zusatznutzens.

3.2 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserhebungen erfasst. In der Studie MOXle Teil 2 traten im Studienverlauf keine Todesfälle auf.

3.3 Morbidität

PGI-C

Die Ergebnisse der Responderanalysen für eine Verbesserung (< 4 Punkte) bzw. Verschlechterung (> 4 Punkte) im PGI-C in der Studie MOXle Teil 2 sind in Tabelle 10 dargestellt. Es zeigten sich keine statistisch signifikanten Unterschiede im PGI-C (Verbesserung oder Verschlechterung) zwischen Omaveloxolon und Placebo zu Woche 48. Im Omaveloxolon-Arm fehlen zu Woche 48 Angaben von 7 Personen (13,7 %) und im Placebo-Arm von 1 Person (1,9 %). Personen mit fehlenden Daten wurden in der Analyse nicht berücksichtigt.

Tabelle 10: Responderanalysen für eine Verbesserung bzw. Verschlechterung im PGI-C zu Woche 48, Studie MOXle Teil 2, ITT

Studie MOXle Teil 2 PGI-C	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)	RR [95%-KI] ¹⁾ ; p-Wert
Verbesserung ²⁾ zu Woche 48	N = 44 19 (43,2)	N = 51 13 (25,5)	1,69 [0,95; 3,01]; 0,08
Verschlechterung ³⁾ zu Woche 48	N = 44 13 (29,5)	N = 51 23 (45,1)	0,66 [0,38; 1,14]; 0,13

¹⁾ Post hoc berechnet: RR mit 95%-KI mittels Cochran-Mantel-Haenszel-Test unter Berücksichtigung des Pescavus-Status.

²⁾ Verbesserung ist definiert als < 4 Punkte (jegliche Verbesserung) im PGI-C.

³⁾ Verschlechterung ist definiert als > 4 Punkte (jegliche Verschlechterung) im PGI-C.

Abkürzungen: ITT: Intention to treat; KI: Konfidenzintervall; PGI-C: Patient Global Impression of Change; RR: Relatives Risiko.

Der pU beschreibt im Nutzendossier, dass sich bei Personen mit einer GAA1-Repeat-Länge ≥ 675 für die Verbesserung im PGI-C ein statistisch signifikanter Vorteil zugunsten von Omaveloxolon im Vergleich zu Placebo zeigt. GAA1-Repeat-Länge ≥ 675 liegt nur für ca. 80 % der ITT-Population vor. Da kein Interaktionstest für die Subgruppenanalyse sowie keine Angaben zu fehlenden Werten vorliegen, sind die Ergebnisse nicht bewertbar.

9-HPT

Es liegen keine Ergebnisse für den Endpunkt „9-HPT“ für die Zeit in Sekunden vor. Da die Ergebnisse für die vom pU dargestellte Geschwindigkeit zu Woche 48 nicht statistisch

signifikant sind, wird davon ausgegangen, dass die Ergebnisse für die Zeit in Sekunden ebenfalls nicht statistisch signifikant sein werden.

T25-FWT

Es liegen keine Ergebnisse für den Endpunkt „T25-FWT“ für die Zeit in Sekunden vor. Da die Ergebnisse für die vom pU dargestellte Geschwindigkeit zu Woche 48 nicht statistisch signifikant sind, wird davon ausgegangen, dass die Ergebnisse für die Zeit in Sekunden ebenfalls nicht statistisch signifikant sein werden.

ADL

Tabelle 11: Veränderung im ADL zu Woche 48, Studie MOXle Teil 2; ITT-Population

Studie MOXle Teil 2 ADL ¹⁾	Omaveloxolon N = 51	Placebo N = 52
Baseline n (%) MW (SD)	51 (100) 11,0 (4,5)	52 (100) 9,9 (4,7)
Veränderung von Baseline zu Woche 48 n (%) LS Mean (SE)	44 (86,3) 0,28 (0,42)	51 (98,1) 1,05 (0,39)
LS-Mean Differenz [95%-KI] ²⁾ ; p-Wert	-0,78 [-1,93; 0,38]; 0,19	

¹⁾ Skala von 0 bis 36 Punkten; höhere Werte korrespondieren mit stärker ausgeprägter Einschränkung in den Alltagsaktivitäten.

²⁾ MMRM mit Baseline-ADL-Wert, Behandlungsgruppe, Erhebungszeitpunkt, Studienzentrum, Vorliegen eines Pes cavus, Interaktion Behandlungsgruppe x Erhebungszeitpunkt und Interaktion Baseline-ADL-Wert x Erhebungszeitpunkt. Keine Imputation von fehlenden Werten.

Abkürzungen: ADL: Activities of Daily Living; ITT: Intention to treat; KI: Konfidenzintervall; LS: Least Squares; MMRM: Mixed Model for Repeated Measurement; MW: Mittelwert; SD: Standardabweichung; SE: Standardfehler.

3.4 Lebensqualität

SF-36

Die Ergebnisse der Responderanalysen für eine Verschlechterung im SF-36 in der Studie MOXle Teil 2 ist in Tabelle 12 dargestellt. Es zeigten sich keine statistisch signifikanten Unterschiede in den Summenscores PCS und MCS zwischen Omaveloxolon und Placebo zu Woche 48. Im Omaveloxolon-Arm fehlen zu Woche 48 Angaben von 7 Personen (13,7 %) und im Placebo-Arm von 1 Person (1,9 %). Personen mit fehlenden Daten wurden in der Analyse nicht berücksichtigt.

Tabelle 12: Responderanalysen für eine Verschlechterung um 15 % im SF-36 zu Woche 48; Studie MOXle Teil 2; ITT-Population

Studie MOXle Teil 2 SF-36	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)	RR [95%-KI] ¹⁾ ; p-Wert
Psychischer Summencore (MCS)²⁾			
Verschlechterung ³⁾ zu Woche 48	N = 44 3 (6,8)	N = 51 3 (5,9)	1,16 [0,25; 5,42]; 0,85

Studie MOXle Teil 2 SF-36	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)	RR [95%-KI] ¹⁾ ; p-Wert
Körperlicher Summencore (PCS)²⁾			
Verschlechterung ³⁾ zu Woche 48	N = 44 3 (6,8)	N = 51 4 (7,8)	0,89 [0,21; 3,71]; 0,87

¹⁾ Post hoc berechnet: RR mit 95%-KI mittels Cochran-Mantel-Haenszel-Test unter Berücksichtigung des Pescavus-Status. Keine Imputation von fehlenden Werten.

²⁾ Höhere Werte korrespondieren mit höherer Lebensqualität.

³⁾ Eine Verschlechterung wurde vom pU als Veränderung von Baseline zu Woche 48 um $\leq -9,4$ Punkte in der PCS und $\leq -9,6$ Punkte in der MCS definiert. Dies entspricht 15 % der Skalenspannweite.

Abkürzungen: ITT: Intention to treat; KI: Konfidenzintervall; MCS: Mental Component Summary; PCS: Physical Component Summary; pU: pharmazeutischer Unternehmer; RR: Relatives Risiko; SF-36: Short-Form-36 Health Survey.

3.5 Sicherheit

In der Studie MOXle Teil 2 wurden alle UE, die nach der ersten Verabreichung der Studienmedikation bis 30 Tage nach der letzten Einnahme der Studienmedikation auftraten, berücksichtigt. Es liegt keine mediane Beobachtungsdauer vor. Die mediane Behandlungsdauer ist mit 336 Tagen im Omaveloxolon-Arm und 337 Tagen im Placebo-Arm jedoch für beide Studienarme vergleichbar.

Der pU legte keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung mit in die UE-Erfassung eingingen.

Tabelle 13: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE; Studie MOXle Teil 2, Sicherheitspopulation

Studie MOXle Teil 2 Personen mit mindestens einem ...	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)	RR [95%-KI] ¹⁾ ; p-Wert
UE (ergänzend dargestellt)	51 (100)	52 (100)	-
Schwere UE	5 (9,8)	0 (0)	11,21 [0,64; 197,67]; 0,10
SUE	5 (9,8)	3 (5,8)	1,70 [0,43; 6,74]; 0,45
UE, das zum Abbruch der Studienmedikation führte	4 (7,8)	2 (3,8)	2,04 [0,39; 10,65]; 0,40

¹⁾ Keine Angaben zum verwendeten statistischen Modell bzw. Testverfahren.

Abkürzungen: KI: Konfidenzintervall; RR: Relatives Risiko; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Unerwünschte Ereignisse

Tabelle 14: UE jeglichen Schweregrads mit Inzidenz $\geq 10\%$ (bzw. mind. 10 Personen) in einem Studienarm in der Studie MOXle Teil 2; Sicherheitspopulation

Studie MOXle Teil 2 MedDRA-Systemorganklasse Preferred Term	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)	RR [95%-KI] ¹⁾ ; p-Wert
Allg. Erkrankungen und Beschwerden am Verabreichungsort	19 (37,3)	17 (32,7)	1,14 [0,67; 1,93]; 0,63
Ermüdung	11 (21,6)	7 (13,5)	1,60 [0,67; 3,81]; 0,29
Erkrankungen der Atemwege, des Brustraums und Mediastinums	20 (39,2)	14 (26,9)	1,46 [0,83; 2,56]; 0,19
Schmerzen im Oropharynx	9 (17,6)	3 (5,8)	3,06 [0,88; 10,66]; 0,08
Husten	6 (11,8)	4 (7,7)	1,53 [0,46; 5,10]; 0,49
Erkrankungen des Gastrointestinaltrakts	34 (66,7)	21 (40,4)	1,65 [1,13; 2,42]; 0,01
Übelkeit	17 (33,3)	7 (13,5)	2,48 [1,12; 5,46]; 0,02
Abdominalschmerz	11 (21,6)	3 (5,8)	3,74 [1,11; 12,62]; 0,03
Diarrhö	10 (19,6)	5 (9,6)	2,04 [0,75; 5,55]; 0,16
Erbrechen	8 (15,7)	6 (11,5)	1,36 [0,51; 3,64]; 0,54
Erkrankungen der Geschlechtsorgane und der Brustdrüse	7 (13,7)	0 (0)	15,29 [0,90; 260,90]; 0,06
Erkrankungen der Haut und des Unterhautgewebes	16 (31,4)	7 (13,5)	2,33 [1,05; 5,19]; 0,04
Erkrankungen des Nervensystems	27 (52,9)	22 (42,3)	1,25 [0,83; 1,89]; 0,28
Kopfschmerzen	19 (37,3)	13 (25,0)	1,49 [0,83; 2,69]; 0,19
Schwindel	3 (5,9)	6 (11,5)	0,51 [0,14; 1,93]; 0,32
Herzerkrankungen	5 (9,8)	7 (13,5)	0,73 [0,25; 2,15]; 0,57
Infektionen und parasitäre Erkrankungen	32 (62,7)	30 (57,7)	1,09 [0,79; 1,49]; 0,60
Infektion der oberen Atemwege	14 (27,5)	15 (28,8)	0,95 [0,51; 1,77]; 0,88
Nasopharyngitis	7 (13,7)	9 (17,3)	0,79 [0,32; 1,97]; 0,62
Psychiatrische Erkrankungen	6 (11,8)	9 (17,3)	0,68 [0,26; 1,77]; 0,43

Studie MOXle Teil 2 MedDRA-Systemorganklasse Preferred Term	Omaveloxolon N = 51 n (%)	Placebo N = 52 n (%)	RR [95%-KI] ¹⁾ ; p-Wert
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	24 (47,1)	21 (40,4)	1,17 [0,75; 1,81]; 0,50
Arthralgie	9 (17,6)	10 (19,2)	0,92 [0,41; 2,07]; 0,84
Muskelspasmen	8 (15,7)	3 (5,8)	2,72 [0,76; 9,68]; 0,12
Rückenschmerzen	7 (13,7)	4 (7,7)	1,78 [0,56; 5,73]; 0,33
Stoffwechsel- und Ernährungsstörungen	8 (15,7)	3 (5,8)	2,72 [0,76; 9,68]; 0,12
Verminderter Appetit	6 (11,8)	2 (3,8)	3,06 [0,65; 14,46]; 0,16
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	33 (64,7)	32 (61,5)	1,05 [0,78; 1,41]; 0,74
Kontusion	17 (33,3)	19 (36,5)	0,91 [0,54; 1,55]; 0,73
Exkoration	13 (25,5)	12 (23,1)	1,10 [0,56; 2,19]; 0,78
Risswunde	8 (15,7)	8 (15,4)	1,02 [0,41; 2,51]; 0,97
Bänderzerrung	5 (9,8)	8 (15,4)	0,64 [0,22; 1,82]; 0,40
Untersuchungen	27 (52,9)	8 (15,4)	3,44 [1,73; 6,85]; < 0,001
Alanin-Aminotransferase erhöht	19 (37,3)	1 (1,9)	19,37 [2,69; 139,39]; 0,003
Aspartat-Aminotransferase erhöht	11 (21,6)	1 (1,9)	11,22 [1,50; 83,74]; 0,02

¹⁾ Keine Angaben zum verwendeten statistischen Modell bzw. Testverfahren.

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

Schwere unerwünschte Ereignisse

Schwere UE traten in keiner SOC und PT bei $\geq 5\%$ der Personen in einem Studienarm auf.

Schwerwiegende unerwünschte Ereignisse

Einzig im SOC „Herzkrankungen“ traten bei $\geq 5\%$ der Personen in einem Studienarm SUE auf. Im Omaveloxolon-Arm traten bei N = 3 (5,9 %) und im Placebo-Arm bei N = 1 (1,9 %) SUE auf. Effektschätzer liegen nicht vor.

UE von besonderem Interesse

Es waren keine UE von besonderem Interesse präspezifiziert.

4 Diskussion der Methodik und Ergebnisse

4.1 Zulassungsstatus und Zulassungspopulation von Omaveloxolon

Omaveloxolon (Skyclarys®) wird zur Behandlung der Friedreich-Ataxie bei Erwachsenen und Jugendlichen ab 16 Jahren angewendet. Gemäß Fachinformation beträgt die zugelassene Dosierung 150 mg Omaveloxolon (3 Hartkapseln zu je 50 mg) einmal täglich [23].

In die für die Nutzenbewertung herangezogene pivotale Studie MOXle Teil 2 wurden Patientinnen und Patienten mit genetisch bestätigter Friedreich-Ataxie und einem mFARS-Score zwischen ≥ 20 und ≤ 80 im Alter von ≥ 16 und ≤ 40 Jahren eingeschlossen. Zudem wurden Personen mit klinisch signifikanten Herzerkrankungen und unkontrolliertem Diabetes mellitus von der Studie ausgeschlossen. Im Anwendungsgebiet besteht eine solche Einschränkung nicht, es lassen sich somit keine Aussagen zu Patientinnen und Patienten im Alter > 40 Jahre und mit einem mFARS von < 20 oder > 80 ableiten. Insbesondere schwer erkrankte Personen sind somit nicht umfasst. Die Dosierung und Behandlungsdauer von Omaveloxolon erfolgte FI-konform.

Die Studienpopulation wurde in Europa (jeweils 1 Studienzentrum in Österreich, Italien und Vereinigtes Königreich), USA (7 Studienzentren) und Australien (1 Studienzentrum) rekrutiert. Von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ist auszugehen.

4.2 Design und Methodik der Studie

Studiendesign

Die MOXle-Studie ist eine multizentrische, randomisierte, kontrollierte, doppelt-verblindete Phase-II-Studie zur Untersuchung der Sicherheit und Wirksamkeit von Omaveloxolon im Vergleich zu Placebo bei Patientinnen und Patienten mit Friedreich-Ataxie im Alter von 16 bis 40 Jahren. Die Studie besteht aus 3 Teilen. Teil 1 ist eine randomisiert-kontrollierte, doppelblinde Dosisfindungsphase über 12 Wochen zur Bestimmung der Omaveloxolon-Dosis für Teil 2. Teil 2 ist eine randomisiert-kontrollierte, doppelblinde Studienphase zur Untersuchung der Wirksamkeit und Sicherheit von Omaveloxolon vs. Placebo über 48 Wochen, mit einer anschließenden 4-wöchigen Nachbeobachtung der Sicherheit. Nach Abschluss von Teil 1 und Teil 2 und Prüfung der Eignung konnten die Studienteilnehmenden in die 144-wöchige, einarmige Langzeit-Extensionsphase (Teil 3: OLE) übergehen. Zum Zeitpunkt der Nutzenbewertung sind Teil 1 und Teil 2 der Studie abgeschlossen; Teil 3 der Studie läuft noch.

Zur Beantwortung der Fragestellung für die Nutzenbewertung wird die doppelblinde RCT-Phase (Teil 2) der Studie MOXle herangezogen. Die Dosierung in Teil 2 erfolgte in der Fachinformation empfohlenen Dosis (150 mg Omaveloxolon). Patientinnen und Patienten wurden 1:1 in die Behandlungsarme randomisiert, stratifiziert nach Pes-cavus-Status (Pes cavus vs. kein Pes cavus). Pes cavus ist definiert als eine muskuloskelettale Fußdeformität, die durch ein hohes Fußgewölbe gekennzeichnet ist (= Hohlfuß). Die Subpopulation mit Pes cavus war auf 20 % der gesamten randomisierten Population begrenzt.

Studienpopulation

In der Studie MOXle Teil 2 wurden 51 Personen in den Omaveloxolon-Arm und 52 in den Placebo-Arm randomisiert (ITT-Population). Im Omaveloxolon-Arm haben 40 Personen und im Placebo-Arm 42 Personen keinen Pes Cavus (FAS). Die FAS-Population ist für die Nutzenbewertung nicht relevant, da die gesamte ITT-Population herangezogen wird. 44

Personen (86,3 %) im Omaveloxolon-Arm und 50 (96,2 %) im Placebo-Arm schlossen die 48-wöchige Behandlung ab. In beiden Behandlungsarmen beendete jeweils 1 Person mehr das Studien-Follow-up bis Woche 52, so dass insgesamt 6 Personen (11,8 %) im Omaveloxolon-Arm und 1 (1,9 %) im Placebo-Arm die Studie abbrachen. 4 Personen (7,8 %) im Omaveloxolon-Arm und 1 Person (1,9 %) im Placebo-Arm zogen die Einverständniserklärung zurück und 2 Personen (3,9 %) im Omaveloxolon-Arm brachen aufgrund von administrativen Gründen die Studie ab.

Die mediane Behandlungsdauer war mit 336 bzw. 337 Tagen vergleichbar zwischen den Behandlungsarmen. Angaben für die mediane Beobachtungsdauer konnten nicht identifiziert werden. Die Patientencharakteristika zu Baseline waren in beiden Behandlungsarmen größtenteils vergleichbar (Tabelle 9). In der Studie MOXle Teil 2 beträgt das mediane Alter im Omaveloxolon-Arm 22,0 Jahre und im Placebo-Arm 21,0 Jahre. Jedoch war die Anzahl an Jugendlichen (≤ 18 Jahre; Einschluss ab 16 Jahre) im Placebo-Arm (28,8 %) fast doppelt so hoch wie im Omaveloxolon-Arm (17,6 %). Nahezu alle Studienteilnehmenden waren europäischer Abstammung. Die Mehrheit der Patientinnen und Patienten war gehfähig (> 90 %) und ca. die Hälfte benötigte noch keine Mobilitätshilfen. 19,6 % bzw. 21,6 % im Omaveloxolon-Arm benutzen Gehhilfen und sitzen im Rollstuhl; im Placebo-Arm liegt der Anteil bei jeweils 15,4 %.

Wesentliche Unterschiede zwischen den Behandlungsarmen zeigen sich in den folgenden Baselinecharakteristika: 60,8 % der Studienteilnehmenden im Omaveloxolon-Arm waren weiblich, im Placebo-Arm waren lediglich 32,7 % weiblich. Zudem wiesen Studienteilnehmende im Omaveloxolon-Arm häufiger eine Kardiomyopathie (49,0 %) sowie operative Behandlung der Skoliose (31,4 %) als im Placebo-Arm (28,8 % bzw. 19,2 %) auf. Daten zur GAA1-Repeat-Länge liegen nur von 41 Personen im Omaveloxolon-Arm und 43 Personen im Placebo-Arm vor. Der Anteil an Personen mit einer GAA1-Repeat-Länge ≥ 675 war jedoch höher im Omaveloxolon-Arm (51,0 %) als im Placebo-Arm (40,4 %). Die Anzahl der GAA-Repeats ist ein prognostischer Faktor und korreliert mit einem früheren Beginn und der Schwere der Friedreich-Erkrankung [5,8]. Kardiomyopathien sind die Haupttodesursachen bei der Friedreich-Ataxie [34]. Es ist nicht auszuschließen, dass im Omaveloxolon-Arm stärker betroffene Personen enthalten sind.

Studienmedikation

Alle Personen der ITT-Population erhielten die vorgesehene Studienmedikation. Diese umfasste im Interventions-Arm 150 mg Omaveloxolon und im Kontroll-Arm Placebo. Laut Modul 4 wird den Patientinnen und Patienten zusätzlich im Placebo-Arm Best Supportive Care (BSC) gegeben. Es ist unklar, was unter BSC verstanden wird, da hierzu Informationen in den Studienunterlagen fehlen.

Verzerrungspotential

Das Verzerrungspotential auf Studienebene wird als hoch bewertet. Gründe dafür sind die Unterschiede zwischen den Behandlungsarmen von > 10 % in den genannten Baselinecharakteristika und vorzeitigem Studienabbruch sowie einer möglichen unbeabsichtigten Entblindung aufgrund des häufigen Auftretens charakteristischer UE in den SOC „Erkrankungen des Gastrointestinaltrakts“ und „Untersuchungen“ im Omaveloxolon-Arm.

Indirekter Vergleich

Der vom pU als supportive Evidenz vorgelegte indirekte Vergleich von Omaveloxolon gegenüber BSC für den primären Endpunkt „mFARS“ wird aufgrund methodischer

Limitationen, welche sich im Wesentlichen auf das Vorgehen bei der Confounderauswahl, die fehlende zeitliche Parallelität, den möglichen Selektionsbias und die nicht adäquate Operationalisierung der mFARS beziehen, nicht in der Nutzenbewertung berücksichtigt (Kapitel 2.6).

4.3 Mortalität

Todesfälle wurden im Rahmen der Sicherheitserhebungen erfasst. Es traten in der Studie MOXle Teil 2 im Studienverlauf keine Todesfälle auf.

4.4 Morbidität

In der Endpunktkategorie „Morbidität“ wurden die patientenrelevanten Endpunkte „PGI-C“, „9-HPT“, „T25-FWT“ und „ADL“ für die Nutzenbewertung herangezogen.

Die Endpunkte „mFARS“, „Maximale Belastbarkeit mittels Peak Work Maximal Exercise Test“, „CGI-C“ und „Häufigkeit von Stürzen“ werden aus unterschiedlichen Gründen nicht für die Nutzenbewertung herangezogen. Die detaillierte Bewertung der Endpunkte ist im Kapitel 2.3 abgebildet.

Der primäre Endpunkt der Studie MOXle Teil 2 zur Erhebung der Krankheitsschwere der Patientinnen und Patienten mit Friedreich-Ataxie ist die mFARS. Da der in der Studie MOXle Teil 2 eingesetzte Fragebogen mit maximal zu erreichender Punktzahl von 99 nicht der Validierung zugrundeliegenden und etablierten Version des Fragebogens mit einer maximal zu erreichenden Punktzahl von 93 entspricht, wurde der Endpunkt nicht berücksichtigt (siehe Kapitel 2.3.2).

Für die beiden Endpunkte „9-HPT“ und „T25-FWT“ liegen nur Ergebnisse für die Geschwindigkeit vor. Zur Beurteilung der Veränderung der Feinmotorik der oberen Gliedmaßen mittels 9-HPT sowie der Beinfunktion mittels T25-FWT wird die Dauer (in Sekunden) als adäquater für eine inhaltlich sinnvolle und verständliche Interpretation angesehen. Da die Ergebnisse für die vom pU dargestellte Geschwindigkeit zu Woche 48 für die Endpunkte 9-HPT und T25-FWT nicht statistisch signifikant sind, wird davon ausgegangen, dass die Ergebnisse für die Zeit in Sekunden ebenfalls nicht statistisch signifikant sein werden.

Für den Endpunkt „PGI-C“ zeigten sich in den Responderanalysen für eine Verbesserung (< 4 Punkte) bzw. Verschlechterung (> 4 Punkte) keine statistisch signifikanten Unterschiede zwischen Omaveloxolon und Placebo zu Woche 48. Im Omaveloxolon-Arm fehlen zu Woche 48 Angaben von 7 Personen (13,7 %) und im Placebo-Arm von 1 Person (1,9 %). Personen mit fehlenden Daten wurden in der Analyse nicht berücksichtigt.

Für den Endpunkt „ADL (Einschränkungen bei Alltagsaktivitäten)“ liegt kein statistisch signifikanter Unterschied zwischen Omaveloxolon und Placebo zu Woche 48 vor. Im Omaveloxolon-Arm fehlen zu Woche 48 Angaben von 7 Personen (13,7 %) und im Placebo-Arm von 1 Person (1,9 %).

Relevante Confounder (siehe Kapitel 2.6 indirekter Vergleich) wie z. B. Geschlecht und GAA1-Repeat-Länge ≥ 675 (nur für PGI-C vorliegend) wurden trotz Unterschieden von > 10 % zwischen den Behandlungsarmen nicht in den statistischen Auswertungen berücksichtigt. Es ist unklar, wie sich das auf das Verzerrungspotential auswirkt. Insgesamt wird das Verzerrungspotential für die Endpunkte „PGI-C“ und „ADL“ aufgrund des Unterschiedes von 11,8 % in den Rücklaufquoten zu Woche 48 zwischen den Behandlungsarmen als hoch eingestuft.

4.5 Lebensqualität

Für den Endpunkt „SF-36“ liegt für die Verschlechterung um 15 % der Skalenspannweite der Summenscores PCS und MCS kein statistisch signifikanter Unterschied zwischen Omaveloxolon und Placebo zu Woche 48 vor. Im Omaveloxolon-Arm fehlen zu Woche 48 Angaben von 7 Personen (13,7 %) und im Placebo-Arm von 1 Person (1,9 %). Personen mit fehlenden Daten wurden in den Analysen nicht berücksichtigt. Aufgrund des Unterschiedes von 11,8 % in den Rücklaufquoten zu Woche 48 zwischen den Behandlungsarmen wird das Verzerrungspotential als hoch eingestuft.

4.6 Sicherheit

In der Studie MOXle Teil 2 entspricht die Sicherheitspopulation der ITT-Population. Es wurden alle UE berücksichtigt, die nach der ersten Verabreichung der Studienmedikation bis 30 Tage nach der letzten Einnahme der Studienmedikation auftraten. Es liegt keine mediane Beobachtungsdauer für UE vor. Die mediane Behandlungsdauer ist mit 336 Tagen im Omaveloxolon-Arm und 337 Tagen im Placebo-Arm jedoch für beide Studienarme vergleichbar.

Der pU legte keine zusätzlichen Auswertungen unter Nichtberücksichtigung von erkrankungsbezogenen Ereignissen bzw. Ereignissen der Grunderkrankung vor. Es ist nicht auszuschließen, dass Ereignisse der Grunderkrankung wie z. B. Herzerkrankungen und Muskelspasmen mit in die UE-Erfassung eingingen.

Zu Woche 52 zeigten sich für schwere UE, SUE sowie UE, die zum Abbruch der Studienmedikation führten, in der Studie MOXle Teil 2 keine statistisch signifikanten Unterschiede zwischen Omaveloxolon und Placebo.

Bei den UE jeglichen Schweregrads nach SOC und PT mit einer Inzidenz $\geq 10\%$ zeigten sich im SOC „Erkrankungen des Gastrointestinaltrakts“ (RR [95%-KI]: 1,65 [1,13; 2,42]; $p = 0,01$) und den dazugehörigen PT „Übelkeit“ (RR [95%-KI]: 2,48 [1,12; 5,46]; $p = 0,02$) und „Abdominalschmerzen“ (RR [95%-KI]: 3,74 [1,11; 12,62]; $p = 0,03$) und im SOC „Erkrankungen der Haut und des Unterhautgewebes“ (RR [95%-KI]: 2,33 [1,05; 5,19], $p = 0,04$) statistisch signifikante Unterschiede zuungunsten von Omaveloxolon gegenüber Placebo. Des Weiteren zeigten sich im SOC „Untersuchungen“ (RR [95%-KI]: 3,44 [1,73; 6,85]; $p \leq 0,001$) und den dazugehörigen PT „Alanin-Aminotransferase erhöht“ (RR [95%-KI]: 19,37 [2,69; 139,39]; $p = 0,003$) und „Aspartat-Aminotransferase erhöht“ (RR [95%-KI]: 11,22 [1,50; 83,74]; $p = 0,02$) statistisch signifikante Unterschiede zuungunsten von Omaveloxolon gegenüber Placebo.

Schwere UE traten in keiner SOC und PT bei $\geq 5\%$ der Personen in einem Studienarm auf. SUE traten einzig im SOC „Herzerkrankungen“ bei $\geq 5\%$ der Personen in einem Studienarm auf. Im Omaveloxolon-Arm traten bei $N = 3$ (5,9 %) und im Placebo-Arm bei $N = 1$ (1,9 %) SUE auf. Ein Relatives Risiko mit 95%-Konfidenzintervall liegt nicht vor. UE von besonderem Interesse waren nicht präspezifiziert.

Das Verzerrungspotential für den Endpunkt „Unerwünschte Ereignisse“ in der Studie MOXle Teil 2 wird aufgrund des Unterschiedes von ca. 10 % im vorzeitigen Abbruch der Studie zwischen den Behandlungsarmen (Omaveloxolon: 11,8 % und Placebo: 1,9 %) als hoch eingestuft. Zudem ist unklar, wie die Effektschätzer und p-Werte berechnet wurden und inwiefern eine adjustierte Auswertung vorgenommen wurde.

Die EMA beschreibt das Sicherheitsprofil von Omaveloxolon im Allgemeinen als gut verträglich. Ein Anstieg der AST- und ALT-Werte ist ein wichtiges Sicherheitsrisiko, das in einer Sicherheitsstudie nach der Markteinführung (PASS) weiter beobachtet werden muss [6]. Wichtige potentielle Sicherheitsbedenken sind medikamenteninduzierte Leberschäden und kongestive Herzinsuffizienz [6].

5 Anforderung an eine qualitätsgesicherte Anwendung

Die Angaben der Fachinformation sind zu berücksichtigen.

6 Zusammenfassung der Nutzenbewertung

Omaveloxolon ist zugelassen für die Behandlung der Friedreich-Ataxie bei Erwachsenen und Jugendlichen ab 16 Jahren. Die Nutzenbewertung von Omaveloxolon basiert auf der zulassungsbegründenden Studie MOXle (Teil 2), einer multizentrischen, randomisierten, kontrollierten, doppelt-verblindeten Phase-II-Studie zur Untersuchung der Sicherheit und Wirksamkeit von Omaveloxolon im Vergleich zu Placebo bei Patientinnen und Patienten mit Friedreich-Ataxie.

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 Omaveloxolon 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 15: Zusammenfassende Darstellung der Ergebnisse der Studie MOXle Teil 2

Studie MOXle Teil 2 Darstellung der Ergebnisse	Omaveloxolon N = 51			Placebo N = 52			Omaveloxolon vs. Placebo	Effekt
Mortalität								
	$N^{2)}$	Personen mit Ereignis, n (%)		$N^{1)}$	Personen mit Ereignis, n (%)		RR [95%-KI]; p-Wert	
Todesfälle²⁾								
	51 ³⁾	0 (0)		52 ³⁾	0 (0)		-	\leftrightarrow
Morbidität								
	$N^{4)}$	Personen mit Ereignis, n (%)		$N^{4)}$	Personen mit Ereignis, n (%)		RR [95%-KI] ⁵⁾ ; p-Wert	
PGI-C								
Verbesserung ⁶⁾ zu Woche 48	44	19 (43,2)		51	13 (25,5)		1,69 [0,95; 3,01]; 0,08	\leftrightarrow
Verschlechterung ⁷⁾ zu Woche 48	44	13 (29,5)		51	23 (45,1)		0,66 [0,38; 1,14]; 0,13	\leftrightarrow
	$N^{4)}$	Baseline MW (SD)	Änderung Woche 48 zu Baseline LS Mean (SE)	$N^{4)}$	Baseline MW (SD)	Änderung Woche 48 zu Baseline LS Mean (SE)	LS-Mean- Differenz [95%-KI] ⁸⁾ ; p-Wert	
ADL⁹⁾								
	51 ¹⁰⁾	11,0 (4,5)	0,28 (0,42)	52 ¹⁰⁾	9,9 (4,7)	1,05 (0,39)	-0,78 [-1,93; 0,38]; 0,19	\leftrightarrow

Studie MOXle Teil 2 Darstellung der Ergebnisse	Omaveloxolon N = 51		Placebo N = 52		Omaveloxolon vs. Placebo	Effekt
Lebensqualität						
	N ⁴⁾	Personen mit Ereignis, n (%)	N ⁴⁾	Personen mit Ereignis, n (%)	RR [95%-KI] ⁵⁾ ; p-Wert	
SF-36 – Psychischer Summencore (MCS)¹¹⁾						
Verschlechterung ¹²⁾ zu Woche 48	44	3 (6,8)	51	3 (5,9)	1,16 [0,25; 5,42]; 0,85	↔
SF-36 – Körperlicher Summencore (PCS)¹¹⁾						
Verschlechterung ¹²⁾ zu Woche 48	44	3 (6,8)	51	4 (7,8)	0,89 [0,21; 3,71]; 0,87	↔
Sicherheit¹³⁾						
	N ¹⁾	Personen mit Ereignis, n (%)	N ¹⁾	Personen mit Ereignis, n (%)	RR [95%-KI] ¹⁴⁾ ; p-Wert	
Schwere UE ¹⁵⁾	51 ³⁾	5 (9,8)	52 ³⁾	0 (0)	11,21 [0,64; 197,67]; 0,10	↔
SUE	51 ³⁾	5 (9,8)	52 ³⁾	3 (5,8)	1,70 [0,43; 6,74]; 0,45	↔
UE, das zum Abbruch der Studienmedikatio n führte	51 ³⁾	4 (7,8)	52 ³⁾	2 (3,8)	2,04 [0,39; 10,65]; 0,40	↔

¹⁾ Sicherheitspopulation (siehe Kapitel 2.4).

²⁾ Todesfälle wurden über die Sicherheit erhoben.

³⁾ 6 Personen (11,8 %) im Omaveloxolon-Arm und 1 Person (1,9 %) im Placebo-Arm brachen die Studie vorzeitig ab.

⁴⁾ ITT-Population (siehe Kapitel 2.4).

⁵⁾ Post hoc berechnet: RR mit 95%-KI mittels Cochran-Mantel-Haenszel-Test unter Berücksichtigung des Pes-cavus-Status.

⁶⁾ Verbesserung ist definiert als < 4 Punkte (jegliche Verbesserung) im PGI-C.

⁷⁾ Verschlechterung ist definiert als > 4 Punkte (jegliche Verschlechterung) im PGI-C.

⁸⁾ MMRM mit Baseline-ADL-Wert, Behandlungsgruppe, Erhebungszeitpunkt, Studienzentrum, Vorliegen eines Pes cavus, Interaktion Behandlungsgruppe x Erhebungszeitpunkt und Interaktion Baseline-ADL-Wert x Erhebungszeitpunkt. Keine Imputation von fehlenden Werten.

⁹⁾ Skala von 0 bis 36 Punkten; höhere Werte korrespondieren mit stärker ausgeprägter Einschränkung in den Alltagsaktivitäten.

¹⁰⁾ Anzahl Personen zu Woche 48: N = 44 im Omaveloxolon-Arm und N = 51 im Placebo-Arm. Die Anzahl entspricht denjenigen Personen, die für die Berechnung der Veränderung von Baseline zu Woche 48 eingingen.

¹¹⁾ Höhere Werte korrespondieren mit höherer Lebensqualität.

¹²⁾ Eine Verschlechterung wurde vom pU als Veränderung von Baseline zu Woche 48 um $\leq -9,4$ Punkte in der PCS und $\leq -9,6$ Punkte in der MCS definiert. Dies entspricht 15 % der Skalenspannweite.

¹³⁾ Weitere Ergebnisse zur Sicherheit (UE nach Organsystemen und Einzelereignisse) sind im Ergebniskapitel 3.5 Sicherheit dargestellt.

¹⁴⁾ Keine Angaben zum verwendeten statistischen Modell bzw. Testverfahren.

¹⁵⁾ Es wurden studieneigene Kriterien zur Schweregradeinteilung verwendet (siehe Kapitel 2.3.4).

Abkürzungen: ADL: Activities of Daily Living; ITT: Intention to treat; KI: Konfidenzintervall; MCS: Mental Component Summary; MMRM: Mixed Model for Repeated Measurement; MW: Mittelwert; PCS: Physical Component Summary; PGI-C: Patient Global Impression of Change; pU: pharmazeutischer Unternehmer; RR: Relatives Risiko; SD: Standardabweichung; SE: Standardfehler; SF-36: Short-Form-36 Health Survey; (S)UE: (Schwerwiegendes) Unerwünschtes Ereignis.

Referenzen

1. **Biogen.** Dossier zur Nutzenbewertung gemäß § 35a SGB V, Modul 4 A: Omaveloxolon (Skyclarys), Behandlung der Friedreich-Ataxie (FA) bei Erwachsenen und Jugendlichen ab 16 Jahren; Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [unveröffentlicht]. 15.03.2024.
2. **Brandsema JF, Stephens D, Hartley J, Yoon G.** Intermediate-dose idebenone and quality of life in Friedreich ataxia. *Pediatr Neurol* 2010;42(5):338-342.
3. **Cutellè C, Rastelli E, Gibellini M, Greco G, Frezza E, Botta A, et al.** Validation of the Nine Hole Peg Test as a measure of dexterity in myotonic dystrophy type 1. *Neuromuscul Disord* 2018;28(11):947-951.
4. **Di Prospero NA, Baker A, Jeffries N, Fischbeck KH.** Neurological effects of high-dose idebenone in patients with Friedreich's ataxia: a randomised, placebo-controlled trial. *Lancet Neurol* 2007;6(10):878-886.
5. **Durr A, Cossee M, Agid Y, Campuzano V, Mignard C, Penet C, et al.** Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N Engl J Med* 1996;335(16):1169-1175.
6. **European Medicines Agency (EMA).** Skyclarys: international non-proprietary name: omaveloxolone; European public assessment report EMEA/H/C/006084/0000 [online]. 14.12.2023. Amsterdam (NED): EMA. [Zugriff: 13.05.2024]. URL: https://www.ema.europa.eu/en/documents/assessment-report/skyclarys-epar-public-assessment-report_en.pdf.
7. **Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, et al.** The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler* 2017;23(5):711-720.
8. **Filla A, De Michele G, Cavalcanti F, Pianese L, Monticelli A, Campanella G, et al.** The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am J Hum Genet* 1996;59(3):554-560.
9. **Hervault M, Balto JM, Hubbard EA, Motl RW.** Reliability, precision, and clinically important change of the Nine-Hole Peg Test in individuals with multiple sclerosis. *Int J Rehabil Res* 2017;40(1):91-93.
10. **Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.** Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-766 e761.
11. **Kieseier BC, Pozzilli C.** Assessing walking disability in multiple sclerosis. *Multiple Sclerosis Journal* 2012;18(7):914-924.
12. **Lynch D.** Correction to safety and efficacy of omaveloxolone in Friedreich ataxia (MOXIe Study). *Ann Neurol* 2023;94(6):1190.
13. **Lynch D, Goldsberry A, Rummey C, Farmer J, Boesch S, delatycki m, et al.** Direct utility of natural history data in analysis of clinical trials: Propensity match-based analysis of omaveloxolone in Friedreich ataxia using the FA-COMS dataset. *medRxiv* 2022:2022.2008. 2012.22278684.

14. **Lynch DR, Chin MP, Delatycki MB, Subramony SH, Corti M, Hoyle JC, et al.** Safety and efficacy of omaveloxolone in Friedreich ataxia (MOXIe Study). *Ann Neurol* 2021;89(2):212-225.
15. **Lynch DR, Farmer JM, Tsou AY, Perlman S, Subramony SH, Gomez CM, et al.** Measuring Friedreich ataxia: complementary features of examination and performance measures. *Neurology* 2006;66(11):1711-1716.
16. **Lynch DR, Farmer JM, Wilson RL, Balcer LJ.** Performance measures in Friedreich ataxia: potential utility as clinical outcome tools. *Mov Disord* 2005;20(7):777-782.
17. **Lynch DR, Goldsberry A, Rummey C, Farmer J, Boesch S, Delatycki MB, et al.** Propensity matched comparison of omaveloxolone treatment to Friedreich ataxia natural history data. *Annals of Clinical and Translational Neurology* 2023;11(1):4-16.
18. **Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al.** Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Multiple Sclerosis Journal* 2017;23(5):704-710.
19. **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.
20. **Norman GR, Stratford P, Regehr G.** Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50(8):869-879.
21. **Patel M, Isaacs CJ, Seyer L, Brigatti K, Gelbard S, Strawser C, et al.** Progression of Friedreich ataxia: quantitative characterization over 5 years. *Annals of clinical and translational neurology* 2016;3(9):684-694.
22. **Proud EL, Miller KJ, Bilney B, Morris ME, McGinley JL.** Construct validity of the 9-Hole Peg Test and Purdue Pegboard Test in people with mild to moderately severe Parkinson's disease. *Physiotherapy* 2020;107:202-208.
23. **Reata Ireland.** Skyclarys 50 mg Hartkapsel [online]. 02.2024. Berlin. [Zugriff: 13.05.2024]. URL: <http://www.fachinfo.de>.
24. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402 (part 2), clinical study report [unveröffentlicht]. 05.11.2020.
25. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402 (part 2), statistical analysis plan version 2.0 [unveröffentlicht]. 18.09.2019.
26. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402 extension: comparison of patients treated with omaveloxolone to propensity score matched natural history patients, statistical analysis plan version 1.0 [unveröffentlicht]. 20.07.2022.
27. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402 post hoc propensity-matched analysis extension and natural history, clinical study report [unveröffentlicht]. 24.03.2022.

28. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402 post hoc propensity-matched analysis extension and natural history, clinical study report addendum 1 [unveröffentlicht]. 26.10.2022.
29. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402, study protocol version 12.0 [unveröffentlicht]. 14.11.2022.
30. **Reata Pharmaceuticals.** A phase 2 study of the safety, efficacy, and pharmacodynamics of RTA 408 in the treatment of Friedreich's ataxia (MOXIe); study 408-C-1402, Zusatzanalysen [unveröffentlicht]. 12.03.2024.
31. **Rummey C, Corben LA, Delatycki MB, Subramony S, Bushara K, Gomez CM, et al.** Psychometric properties of the Friedreich ataxia rating scale. *Neurology Genetics* 2019;5(6):371.
32. **Rummey C, Zesiewicz TA, Perez-Lloret S, Farmer JM, Pandolfo M, Lynch DR.** Test-retest reliability of the Friedreich's ataxia rating scale. *Annals of Clinical and Translational Neurology* 2020;7(9):1708-1712.
33. **Subramony S, May W, Lynch D, Gomez C, Fischbeck K, Hallett M, et al.** Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. *Neurology* 2005;64(7):1261-1262.
34. **Tsou AY, Paulsen EK, Lagedrost SJ, Perlman SL, Mathews KD, Wilmot GR, et al.** Mortality in Friedreich ataxia. *J Neurol Sci* 2011;307(1-2):46-49.
35. **Ware Jr JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical care* 1992:473-483.