

Anwendungsbegleitende Datenerhebung zu Risdiplom: Prüfung des Studienprotokolls (Version 3.0) und des statistischen Analyseplans (Version 3.0)

3. Addendum zum Projekt A21-131

A horizontal bar composed of 18 rectangular segments of varying shades of blue and grey. The word 'ADDENDUM' is written in white capital letters on a dark blue segment that spans across the middle of the bar.

ADDENDUM

Projekt: A24-57

Version: 1.0

Stand: 16.07.2024

IQWiG-Berichte – Nr. 1826

DOI: 10.60584/A24-57

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Anwendungsbegleitende Datenerhebung zu Risdiplom: Prüfung des Studienprotokolls (Version 3.0) und des statistischen Analyseplans (Version 3.0) – 3. Addendum zum Projekt A21-131

Auftraggeber

Gemeinsamer Bundesausschuss

Datum des Auftrags

06.05.2024

Interne Projektnummer

A24-57

DOI-URL

<https://dx.doi.org/10.60584/A24-57>

Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8
50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

Zitiervorschlag

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Anwendungsbegleitende Datenerhebung zu Risdiplom: Prüfung des Studienprotokolls (Version 3.0) und des statistischen Analyseplans (Version 3.0); 3. Addendum zum Projekt A21-131 [online]. 2024 [Zugriff: TT.MM.JJJJ]. URL: <https://dx.doi.org/10.60584/A24-57>.

Schlagwörter

Risdiplom, Muskelatrophie – Spinale, Nutzenbewertung, Register, Forschungsdesign, Peer Review – Forschung

Keywords

Risdiplom, Muscular Atrophy – Spinal, Benefit Assessment, Registries, Research Design, Peer Review – Research

An dem Addendum beteiligte Mitarbeiterinnen und Mitarbeiter des IQWiG

- Anna-Katharina Barnert
- Lars Beckmann
- Tatjana Hermanns
- Katharina Hirsch
- Simone Johner
- Volker Vervölgyi

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis.....	vi
1 Hintergrund.....	1
2 Prüfung der Unterlagen zur Planung der AbD von Risdiplom.....	2
2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen	2
2.2 Anmerkungen zur Umsetzung der Auflagen des G-BA durch den pU	2
2.2.1 PICO: Patientenpopulation.....	2
2.2.2 Fragestellung gemäß PICO: Outcome	3
2.2.3 Fragestellung gemäß PICO: Outcome, Morbidität.....	4
2.2.4 Fragestellung gemäß PICO: Outcome, Bayley III	5
2.2.5 Fragestellung gemäß PICO: Outcome, CHOP-INTEND	6
2.2.6 Fragestellung gemäß PICO: Outcome, 6-Minuten-Gehstrecke.....	6
2.2.7 Fragestellung gemäß PICO: Outcome, Erreichen motorischer Meilensteine	7
2.2.8 Fragestellung gemäß PICO: Outcome, bulbäre Funktion.....	8
2.2.9 Fragestellung gemäß PICO: Outcome, Weitere Komplikationen der Erkrankung.....	9
2.2.10 Fragestellung gemäß PICO: Outcome, Unerwünschte Ereignisse (UE).....	9
2.2.11 Fragestellung gemäß PICO: Outcome, Schwerwiegende unerwünschte Ereignisse (SUE)	10
2.2.12 Datenquelle / Studiendesign: Allgemein.....	10
2.2.13 Datenquelle: Einbindung weiterer Register	11
2.2.14 Datenquelle: Confounder	11
2.2.15 Datenquelle: Meldezeitpunkte.....	13
2.2.16 Datenquelle: Vollständigkeit der Daten	14
2.2.17 Studiendesign: Indexdatum.....	15
2.2.18 Studiendesign: Zuordnung zu den Behandlungsgruppen	15
2.2.19 Studiendesign: Fallzahlplanung.....	16
2.2.20 Studiendesign: Abbruchkriterien	17
2.2.21 Studiendesign: Beobachtungsdauer.....	18
2.2.22 Studiendesign: Angaben zum Verlauf der Datenerhebung	19
2.2.23 Auswertung der Daten: Estimand	19
2.2.24 Auswertung der Daten: Responderanalysen	21
2.2.25 Auswertung der Daten: stetige Auswertungen.....	21

2.2.26	Auswertung der Daten: Sensitivitätsanalysen	22
2.2.27	Auswertung der Daten: Subgruppenanalysen	22
2.2.28	Auswertung der Daten: Propensity-Score-Verfahren	24
2.2.29	Auswertung der Daten: unerwünschte Ereignisse (UE)	24
2.2.30	Auswertung der Daten: Umgang mit fehlenden Werten.....	25
2.2.31	Auswertung der Daten: Fehlende Angaben	26
3	Literatur	28
Anhang A	Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern ..	31

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AbD	Anwendungsbegleitende Datenerhebung
CHOP-INTEND	Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders
CTCAE	Common Terminology Criteria for Adverse Events
eCRF	Elektronischer Erfassungsbogen
G-BA	Gemeinsamer Bundesausschuss
HFMSE	Hammersmith Functional Motor Scale Expanded
HR	Hazard Ratio
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
MMRM	Mixed Model Repeated measures
PICO	Patient, Intervention, Comparator, Outcome
pU	pharmazeutischer Unternehmer
RULM	Revised Upper Limb Module
SAP	statistischer Analyseplan
SDV	Source Data Verification
SGB	Sozialgesetzbuch
SMN	Survival Motor Neuron
SUE	schwerwiegendes unerwünschtes Ereignis
UE	unerwünschtes Ereignis

1 Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) am 06.05.2024 mit der Prüfung des Studienprotokolls und des statistischen Analyseplans (SAP) zur anwendungsbegleitenden Datenerhebung (AbD) zu Risdiplam beauftragt.

In seiner Sitzung am 21.07.2022 hat der G-BA beschlossen, eine AbD und Auswertungen nach § 35a Absatz 3b Satz 1 SGB V für den Wirkstoff Risdiplam in der Indikation „spinale Muskelatrophie“ zu fordern [1,2]. Dem Beschluss liegt unter anderem das vom IQWiG erstellte Konzept für eine AbD zu Risdiplam zugrunde (Rapid Report 21-131 vom 15.02.2022) [3].

Zur Prüfung, ob die Anforderungen des G-BA an die AbD und an Auswertungen durch die vom pharmazeutischen Unternehmer (pU) erstellten Unterlagen zum Studienprotokoll und SAP umgesetzt worden sind, hat der G-BA dem IQWiG diese Unterlagen übermittelt [4,5] und mit der Prüfung dieser Unterlagen beauftragt.

Auf Basis der Ergebnisse dieser Prüfung (Addendum A23-87 zum Konzept A21-131 [6]) hat der G-BA mit Beschluss vom 04.04.2024 [7,8] formuliert, bei welchen Angaben im vorgelegten Studienprotokoll und SAP Anpassungsbedarf besteht. Der pU hat am 02.05.2024 und 01.07.2024 überarbeitete Unterlagen beim G-BA eingereicht [9,10]. Der G-BA hat dem IQWiG diese Unterlagen am 03.05.2024 und 01.07.2024 übermittelt.

2 Prüfung der Unterlagen zur Planung der AbD von Risdiplam

2.1 Allgemeine Anmerkungen zu den vom pU vorgelegten Unterlagen

Der pU hat das Studienprotokoll und den SAP auf Basis der Anforderungen des G-BA Beschlusses vom 04.04.2024 überarbeitet.

In den vom pU überarbeiteten Dokumenten fehlt eine systematische Auflistung der Anpassungen im Studienprotokoll und SAP entsprechend der vom G-BA geforderten Anpassungen ebenso wie eine generelle Übersicht zur Änderungshistorie.

Die vorliegende Prüfung des Studienprotokolls und SAP bezieht sich ausschließlich auf die vom G-BA genannten noch vorzunehmenden Anpassungen.

2.2 Anmerkungen zur Umsetzung der Auflagen des G-BA durch den pU

2.2.1 PICO: Patientenpopulation

Auflage des G-BA

Die Ein- und Ausschlusskriterien sind um die im G-BA Beschluss genannten Anforderungen an die Datenqualität zu ergänzen.

Das Einschlusskriterium „Visite, die als Baselinevisite definiert wird, muss zwischen 6 Wochen vor und 3 Wochen nach der ersten Applikation der Survival Motor Neuron (SMA)-Medikation liegen“ ist zu streichen.

Bezüglich des Ausschlusskriteriums „Vorbehandlung der Patientinnen und Patienten mit Risdiplam, Nusinersen oder Onasemnogen-Abeparvovec vor Einschluss in das Register“ ist sicherzustellen, dass nicht auch Patientinnen und Patienten ausgeschlossen werden, die lediglich eine Vorbehandlung mit Risdiplam oder Nusinersen in Form einer Brückentherapie erhalten haben.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im Studienprotokoll sind unter den Selektionskriterien Anforderungen an die Datenqualität entsprechend den im Beschluss des G-BA zur Forderung einer AbD [2] genannten Qualitätskriterien ergänzt.

Das Einschlusskriterium „Visite, die als Baselinevisite definiert wird, muss zwischen 6 Wochen vor und 3 Wochen nach der ersten Applikation der SMA-Medikation liegen“ wurde gestrichen.

Die Ausschlusskriterien wurden in Version 3.0 des Studienprotokolls um den Hinweis ergänzt, dass eine initiale Behandlung mit Risdiplam oder Nusinersen für weniger als 3 Monate (im Sinne einer Brückentherapie) nicht zum Ausschluss von Patientinnen oder Patienten führt. Diese Patientinnen und Patienten werden der nachfolgenden Therapie zugeteilt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind sachgerecht.

Unbenommen von der Streichung des Einschlusskriteriums hinsichtlich des Zeitpunkts der Baselinevisite, ist eine eindeutige Beschreibung in den Studienunterlagen zu ergänzen, dass alle Baselinecharakteristika zum Indexdatum erhoben werden müssen (siehe Abschnitt 2.2.17).

Mit Version 3.0 des Studienprotokolls legt der pU eine klare Definition, wann eine Therapie als Brückentherapie gewertet wird, vor. Er setzt hierfür eine Grenze von 3 Monaten an, als Referenz nennt der pU hierfür die Publikation von Proud 2023 [11]. Auch wenn in der Publikation lediglich für eine Vortherapie mit Risdiplam eine klare Grenze von 3 Monaten genannt wird (für Nusinersen wird die Definition hingegen an die Aufsättigungsphase geknüpft), ist das Vorgehen des pU im Sinne eines pragmatischen Ansatzes nachvollziehbar. In der zeitlich parallel im SMARtCARE Register laufenden AbD zu Onasemnogen-Abeparvovec wird abweichend dazu eine bis zu 2-wöchige Behandlung mit Risdiplam vor der Behandlung mit Onasemnogen-Abeparvovec als Brückentherapie aufgefasst.

2.2.2 Fragestellung gemäß PICO: Outcome

Auflage des G-BA

Sofern bei Endpunkten Auswertungen zu mehreren Zeitpunkten geplant sind, ist immer die Auswertung, die den längsten möglichen Beobachtungszeitraum berücksichtigt, als primäre Analyse darzustellen.

Zwecks Übersicht zu den geplanten Erhebungszeitpunkten im Beobachtungsverlauf ist ein Erhebungsplan zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im Studienprotokoll und SAP hat der pU für die betreffenden primären Endpunkte ergänzt, dass diese zu Monat 36 ausgewertet werden sollen. Für die sekundären Endpunkte, für die Auswertungen zu mehreren Zeitpunkten geplant sind, macht er keine Angaben dazu, dass die primäre Analyse zu Monat 36 erfolgt.

Der pU legt mit den Studienunterlagen zwar die allgemeinen Empfehlungen zur Verlaufsbeobachtung von Kindern und Erwachsenen des SMARtCARE-Registers vor, einen Erhebungsplan für die Erhebung der relevanten Daten im Rahmen der AbD fehlt im Studienprotokoll jedoch weiterhin.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind nur teilweise sachgerecht. Für Endpunkte, zu denen zu mehreren Zeitpunkten Auswertungen geplant sind, ist für die Nutzenbewertung jeweils die Auswertung mit dem längsten Beobachtungszeitraum die primäre relevante Analyse.

Durch die konkretisierten Angaben zur Erhebung der einzelnen Endpunkte in den Studienunterlagen ist zwar nachvollziehbar, wann diese erhoben werden sollen. Zur besseren Übersicht wäre in der vorliegenden Situation, in der eine Vielzahl unterschiedlicher Endpunkte erhoben wird, jedoch ein Erhebungsplan zur besseren Übersicht weiterhin empfehlenswert.

2.2.3 Fragestellung gemäß PICO: Outcome, Morbidität

Auflage des G-BA

Um die Vielzahl der Endpunkte zur Beschreibung der motorischen Funktion zu verringern, sind im Studienprotokoll prä-spezifiziert relevante Endpunkte auszuwählen und die Endpunkte zu hierarchisieren.

Der altersgeeignete Einsatz von Instrumenten zur Erhebung der motorischen Funktion ist, insbesondere bei der geplanten Verwendung von Hammersmith Functional Motor Scale Expanded (HFMSSE) und Revised Upper Limb Module (RULM) bei Patientinnen und Patienten mit Typ-2- und Typ-3-SMA, zu prüfen. Im Studienprotokoll ist insbesondere für Operationalisierungen, in denen Veränderungen gegenüber Baseline erhoben werden, zu spezifizieren und sicherzustellen, dass altersgeeignete Instrumente zur Erhebung der motorischen Funktion eingesetzt werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Von den Endpunkten zur Beschreibung der motorischen Funktion wurden die Auswertungen der motorischen Skala des Bayley III gestrichen.

Eine Hierarchisierung der Endpunkte wurde nicht vorgenommen.

Für die Endpunkte HFMSSE und RULM wird spezifiziert, dass entsprechend der Erhebungspraxis im SMARtCARE-Register diese nur bei Patientinnen und Patienten > 2 Jahren eingesetzt werden (der HFMSSE zusätzlich nur bei sitzfähigen Patientinnen und Patienten). Die Auswertungen zur Veränderung im Vergleich zu Baseline für diese Endpunkte sollen nur Patientinnen und Patienten einschließen, für die ein Baselinewert vorliegt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind nicht ausreichend. Es wird weiterhin eine Vielzahl von Endpunkten zur Beschreibung der motorischen Funktion erhoben, ohne deren Auswertung zu hierarchisieren.

Für die Endpunkte HFMSE und RULM (die als Endpunkte für Patientinnen mit Typ-2-SMA bzw. Typ-3-SMA erhoben werden sollen) ist nun zwar der altersgerechte Einsatz beschrieben, allerdings führt dies insbesondere für Patientinnen und Patienten mit Typ-2-SMA (bei denen sich die Erkrankung in der Regel in einem Alter zwischen 6 und 15 Lebensmonaten manifestiert) absehbar dazu, dass für einen relevanten Anteil der untersuchten Patientinnen und Patienten dieser Fragestellung mindestens keine Baselineerhebung durchgeführt werden kann. Zwar schränkt der pU die geplanten Auswertungen dieser Endpunkte auf Patientinnen und Patienten mit einer Baselineerhebung ein, solche Auswertungen schließen damit aber von vorneherein einen relevanten Anteil der Patientinnen und Patienten in den entsprechenden Fragestellungen aus, nämlich diejenigen, die zu Behandlungsbeginn ≤ 2 Jahre waren. Dementsprechend erlauben die Auswertungen keine Aussagen für die jeweils gesamte Population in diesen Fragestellungen. Dies ist nicht sachgerecht.

Auch wenn für die Nutzenbewertung grundsätzlich Auswertungen im Vergleich zu Baseline zu bevorzugen sind, kann es in der vorliegenden Situation sinnvoll sein, eine Auswertung der Endpunkte zu Monat 36 nach Behandlungsbeginn ohne Berücksichtigung der Baselinewerte – ähnlich zur vom pU geplanten Auswertung für die Gehstrecke – vorzunehmen. Das damit einhergehende Verzerrungspotenzial ist bei der Interpretation der Ergebnisse zu berücksichtigen. Darüber hinaus ist zu beachten, dass auch am Ende der Beobachtungsdauer zu Monat 36 der altersgerechte Einsatz der Instrumente gewährleistet ist, um aussagekräftige Ergebnisse zur motorischen Funktion zu erzielen.

Andernfalls ist insbesondere vor dem Hintergrund der Vielzahl an weiteren motorischen Endpunkten die Streichung dieser Endpunkte (HFMSE und RULM) empfehlenswert.

2.2.4 Fragestellung gemäß PICO: Outcome, Bayley III

Auflage des G-BA

Es ist in den Studienunterlagen zu ergänzen, welcher Score (scaled scores, composite scores, percentile ranks, and growth scores) ausgewertet wird. In den Tragenden Gründen spezifiziert der G-BA, dass er sich hierbei auf die Subskala zu den grobmotorischen Fähigkeiten des Bayley III bezieht.

Für alle Operationalisierungen ist der Auswertungszeitpunkt abhängig von der Zeit seit Behandlungsbeginn festzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In der 1. Version des Studienprotokolls und SAP wurde die Erhebung der motorischen Funktion für präsymptomatische Patientinnen und Patienten sowie für Patientinnen und Patienten mit Typ-1-SMA über die Subskala zur grobmotorischen Funktion des Bayley III geplant. Dieser Endpunkt wurde gestrichen. Der Bayley III wird jedoch weiterhin zur Erhebung der bulbären Funktion über die Subskalen expressive und rezeptive Sprache eingesetzt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Durch die Streichung der Erhebung der grobmotorischen Funktion über den Bayley III entfällt die Prüfung der Auflage des G-BA bezogen auf diesen Endpunkt. Punkte, die für die Subskalen expressive und rezeptive Sprache von Relevanz sind, werden in Abschnitt 2.2.8 adressiert.

2.2.5 Fragestellung gemäß PICO: Outcome, CHOP-INTEND

Auflage des G-BA

Bezüglich der Auswertungen zum Anteil der Patientinnen und Patienten mit einem Score von ≥ 40 beim Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP-INTEND) ist in den Studienunterlagen zu ergänzen, inwieweit ein Score von 40 eine geeignete Responseschwelle darstellt. Zudem ist bei der gewählten Operationalisierung zu beachten, dass der gesamte Beobachtungszeitraum abgedeckt ist.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die ursprünglich geplanten Auswertungen zum Anteil der Patientinnen und Patienten mit einem Score von ≥ 40 zum Alter von 12 Monaten sind gestrichen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Da die oben genannten Auswertungen gestrichen wurden, erübrigen sich die vom G-BA genannten Auflagen. Der CHOP-INTEND wird somit ausschließlich als Änderung gegenüber Baseline ausgewertet.

2.2.6 Fragestellung gemäß PICO: Outcome, 6-Minuten-Gehstrecke

Auflage des G-BA

Es ist darzulegen, dass es sich bei der gewählten Responseschwelle (Verbesserung bzw. die Verschlechterung der zurückgelegten Gehstrecke um > 30 m) um eine klinisch relevante Änderung handelt, andernfalls sind stetige Analysen festzulegen.

Sofern der Endpunkt für die Nutzenbewertung herangezogen werden soll, ist eine Auswertung der Gehstrecke zu Monat 36 nach Behandlungsbeginn ohne Berücksichtigung der Baselinewerte festzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU hat die Responseschwelle von > 30 m gestrichen und plant nun für gehfähige Patientinnen und Patienten zum einen die Veränderung gegenüber Baseline als Analyse, zum anderen wurde die Analyse über die Gehstrecke zu Monat 36 ohne Berücksichtigung der Baselinewerte ergänzt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Der pU hat die Anforderungen des G-BA entsprechend umgesetzt. Für die Nutzenbewertung ist die Auswertung im Vergleich zu Baseline die primär relevante Analyse. Im Fall, dass für einen relevanten Anteil von Patientinnen und Patienten Daten zu Baseline fehlen, kann eine Auswertung ohne Berücksichtigung von Baseline sinnvoll sein. Das damit einhergehende Verzerrungspotenzial ist bei der Interpretation der Ergebnisse zu berücksichtigen.

2.2.7 Fragestellung gemäß PICO: Outcome, Erreichen motorischer Meilensteine

Auflage des G-BA

Der Endpunkt „Stehen ohne Unterstützung“ und ein Endpunkt zum Erhalt der motorischen Funktion sind zu ergänzen.

Bezüglich der Operationalisierung des Anteils der Patientinnen und Patienten, die einen Meilenstein zu einem altersgerechten Zeitpunkt erreichen, sind Ereigniszeitanalysen oder Analysen zum Anteil der Patientinnen und Patienten, die zu einem zu definierenden Zeitpunkt (z.B. 12 Monate, 24 Monate und 36 Monate nach Behandlungsstart) einen motorischen Meilenstein erreicht haben, durchzuführen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der Endpunkt „Stehen ohne Unterstützung“ ist in den Studienunterlagen für präsymptomatische Patientinnen und Patienten und Patientinnen und Patienten mit Typ-1-SMA ergänzt.

Für alle 4 Auswertungspopulationen ist ein Endpunkt zum Erhalt der motorischen Funktion in der Operationalisierung „Zeit vom Erreichen des motorischen Meilensteins bis zum dauerhaften Verlust der Meilenstein-Fähigkeit“ ergänzt. Für präsymptomatische Patientinnen und Patienten ebenso wie für Typ-1-SMA Patientinnen und Patienten plant der pU den Erhalt der 3 Meilensteine Sitzen, Stehen und Gehen zu erheben, für Patientinnen und Patienten mit Typ-2-SMA bzw. Typ-3-SMA plant er lediglich den Erhalt des Meilensteins Gehen.

Die Auswertungen zu den motorischen Meilensteinen wurden alle in Ereigniszeitanalysen (Zeit von der ersten Behandlung bis Erreichen des motorischen Meilensteins) geändert.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind weitgehend sachgerecht.

Für Analysen zur Zeit bis zum Erreichen eines motorischen Meilensteines ist festzulegen, dass als Beobachtungsstart das Indexdatum herangezogen werden soll.

Für Patientinnen und Patienten mit Typ-2-SMA und Typ-3-SMA sollte der Erhalt der Meilensteine Sitzen und Stehen ergänzt werden.

2.2.8 Fragestellung gemäß PICO: Outcome, bulbäre Funktion

Auflage des G-BA

Die Operationalisierung des Anteils der Patientinnen und Patienten, die altersgerechte Scores, definiert als Altersäquivalente, in den Subskalen expressive Sprache und rezeptive Sprache zeigen, ist anzupassen, z. B. als Veränderung im Vergleich zu Baseline.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU ergänzt in den Studienunterlagen, dass der Bayley III entsprechend der Richtlinien zur Verlaufsbeobachtung des SMARtCARE-Registers lediglich einmalig im Alter von 24 Monaten erhoben wird.

In den Studienunterlagen beschreibt der pU Responderanalysen zum Anteil der Patientinnen und Patienten, die altersgerechte Scores in den Subskalen expressive Sprache und rezeptive Sprache im Alter von 24 Monaten erreichen, auszuwerten. Im SAP beschreibt er, Patientinnen und Patienten, die keinen altersgerechten Score zu Monat 24 erreichen oder für die zu Monat 24 kein Wert vorliegt (obwohl die Patientin oder der Patient im Alter von 24 Monaten in der Studie war), als Non-Responder zu werten. Als altersgerechter Score wird – entgegen der 1. Version des SAP – ein Growth Score, der über dem 5. Perzentilrang von gesunden Kindern liegt, definiert.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind nicht sachgerecht. Der pU verweist bei der Operationalisierung des Bayley III auf das zugehörige Manual [12]. In diesem sind Perzentilkurven für die beiden Subskalen zur Sprachentwicklung im Altersverlauf von gesunden Kindern dargestellt. Jedoch bleibt offen, inwiefern die 5. Perzentile eine inhaltlich sinnvolle Responseschwelle für die Bewertung der sprachlichen Entwicklung von SMA-Patientinnen und -Patienten darstellt. Dies sollte vom pU begründet werden. Darüber hinaus ist die oben beschriebene, vom pU geplante, Ersetzung fehlender Werten als Non-Responder nicht sachgerecht.

Vor dem Hintergrund der bereits erhobenen Endpunkte zur bulbären Funktion (Schluckfähigkeit und Notwendigkeit nicht-oraler Ernährungsunterstützung), ist die Erhebung dieser Endpunkte verzichtbar. Falls der Endpunkt weiterhin erhoben werden soll, ist eine Erhebung über den gesamten Beobachtungszeitraum und – wie bereits in der 1. Prüfung des Studienprotokolls und SAP beschrieben [6] – für alle Patientinnen und Patienten sicherzustellen. Die einmalige Erhebung im Alter von 24 Monaten ist für die Fragestellung der AbD nicht sachgerecht.

2.2.9 Fragestellung gemäß PICO: Outcome, Weitere Komplikationen der Erkrankung

Auflage des G-BA

Es sind altersgerechte, valide Instrumente zur Erhebung der Symptome Schmerzen oder Fatigue anzuwenden, sofern diese vorliegen. Eine entsprechende Prüfung ist vorzunehmen.

Eine Begründung, warum die Erfassung der Skoliose auf das Auftreten schwerer Ereignisse einschränkt ist, ist zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die Endpunkte Schmerzen und Fatigue sind aus den Studienunterlagen gestrichen.

Die Erfassung der Skoliose ist nicht mehr auf das Auftreten schwerer Ereignisse eingeschränkt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind nachvollziehbar.

2.2.10 Fragestellung gemäß PICO: Outcome, Unerwünschte Ereignisse (UE)

Auflage des G-BA

Bei vergleichbarer Beobachtungsdauer sind Auswertungen zum Anteil der Patientinnen und Patienten mit Ereignis bei den UE-Endpunkten durchzuführen.

Der Endpunkt „Anzahl ungeplanter Hospitalisierungen“ ist verzichtbar und zu streichen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU hat bei den geplanten Auswertungen der UE-Endpunkte die Ereigniszeitanalysen durch Auswertungen zum Anteil der Patientinnen und Patienten mit Ereignis ersetzt.

Der Endpunkt zur Anzahl der UEs, die zu einer ungeplanten Hospitalisierung führen, wurde angepasst und bezieht sich nicht mehr auf die ungeplanten, sondern auf jegliche Hospitalisierungen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vorgenommenen Änderungen sind weitgehend sachgerecht. Der Endpunkt „Anzahl an UEs, die zu einer Hospitalisierung führen“ ist für die Nutzenbewertung weiterhin nicht relevant und daher verzichtbar.

2.2.11 Fragestellung gemäß PICO: Outcome, Schwerwiegende unerwünschte Ereignisse (SUE)

Auflage des G-BA

Eine Beschreibung der Operationalisierung der SUEs ist zu ergänzen. Es ist klarzustellen, dass eine Annäherung über UEs, die zur Hospitalisierung oder einer Verlängerung der Hospitalisierung führen bzw. UEs, die zum Tod führen als Operationalisierung gewählt werden. Die entsprechenden Dokumentationsfelder im SMARtCARE-Register sind zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im Studienprotokoll wurde ergänzt, dass im SMARtCARE-Register keine spezifische Variable für die Erfassung von SUEs existiert. Für die vorliegende AbD sollen SUEs näherungsweise über UEs, die zu einer ungeplanten Hospitalisierung führen oder eine Hospitalisierung verlängern, erfasst werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommene Änderung ist nicht sachgerecht. In der vom pU gewählten Operationalisierung fehlt die Komponente UEs, die zum Tod führen. Zwar werden diese nicht direkt im SMARtCARE-Register erhoben, die Angaben im Freitextfeld der Variable „Todesursache“ können aber – analog zur AbD zu Onasemnogen-Abeparvovec – herangezogen werden, um einen Todesfall durch ein UE bedingt zu klassifizieren. Die Beschreibung der Operationalisierung ist in den Studienunterlagen entsprechend anzupassen. Ebenso sind die entsprechenden Dokumentationsfelder des SMARtCARE-Registers zu ergänzen.

2.2.12 Datenquelle / Studiendesign: Allgemein

Auflage des G-BA

Für Endpunkte, Einschlusskriterien und Confounder sind alle relevanten Datenfelder mit der entsprechenden Operationalisierung im Register in den Studienunterlagen anzugeben. Zudem ist die aktuelle Version des Kodierhandbuchs zur ergänzenden Information vorzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die aktuelle Version des Kodierhandbuchs wurde mit den Studienunterlagen eingereicht.

Für die Endpunkte und Confounder führt der pU im Studienprotokoll die entsprechenden Datenfelder im SMARtCARE-Register auf. Für die Einschlusskriterien wird im Studienprotokoll weiterhin nur die relevante Seite des elektronischen Erfassungsbogens (eCRF) benannt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Für die Einschlusskriterien sind analog zu den Endpunkten und Confoundern die relevanten Datenfelder mit ihrer Operationalisierung im Studienprotokoll zu ergänzen. Diese Angaben sind insbesondere für die Studienmonitore zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung relevant.

Ungeachtet der notwendigen Dokumentation in den Studienunterlagen ist im Abgleich mit dem vom pU vorgelegten Kodierhandbuch des SMARtCARE-Registers davon auszugehen, dass eine Abgrenzung der für die AbD relevanten Patientenpopulationen möglich ist.

2.2.13 Datenquelle: Einbindung weiterer Register

Auflage des G-BA

Sofern weitere Register eingebunden werden sollen, ist die Einbindung der jeweiligen Register vorab zu benennen und die Eignung der jeweiligen Register entsprechend der vom G-BA im Forderungsbeschluss genannten Kriterien zu prüfen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im Studienprotokoll wurde der Satz gestrichen, dass ggf. weitere Register eingebunden werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Aufgrund der vorgenommenen Streichung erübrigen sich die vom G-BA genannten Auflagen.

2.2.14 Datenquelle: Confounder

Auflage des G-BA

Für Patientinnen und Patienten mit Typ 3 SMA ist eine systematische Literaturrecherche zur Identifikation möglicher weiterer potenzieller Confounder erforderlich.

Zudem ist zu beschreiben, auf welche Weise die ausgewählten Confounder in das Modell zur Propensity Score Berechnung eingehen (z. B. stetig oder dichotom) und welchen Einfluss die Einteilung in „less important“ und „very important“ auf die Modellierung des Propensity Scores hat.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU legt einen Bericht über eine systematische Literaturrecherche nach potenziell relevanten Confoundern für das gesamte relevante Anwendungsgebiet der vorliegenden AbD

vor. Für die Validierung der Confounder bindet der pU klinische Expertinnen und Experten ein, die die identifizierten potentiellen Confounder in 3 Gruppen kategorisieren sollten (sehr wichtig, weniger wichtig, nicht wichtig).

Für die sich anschließende Confounder-Auswahl beschreibt der pU folgendes Vorgehen: Potenzielle Confounder, die

- als sehr wichtig eingestuft wurden, werden berücksichtigt.
- als weniger wichtig eingestuft wurden, werden berücksichtigt, wenn sie vom IQWiG gefordert oder in der AbD von Onasemnogen-Abeparvovec gemäß Studienprotokoll berücksichtigt wurden.
- als nicht wichtig kategorisiert wurden, werden nicht berücksichtigt.

Wenn für einen potenziellen Confounder (z. B. motorische oder respiratorische Funktion) mehr als 1 Operationalisierung identifiziert wurde, wird jene, die von den klinischen Expertinnen und Experten als am wichtigsten eingeschätzt wurde, berücksichtigt. Für den Fall, dass mehrere Operationalisierungen als gleich wichtig benannt wurden, erfolgt die Auswahl analog zum Studienprotokoll zur AbD von Onasemnogen-Abeparvovec.

Der pU beschreibt in Tabelle 9 des Studienprotokolls sowie in Tabelle 5 des SAP, welches Skalenniveau (diskret bzw. stetig) die Variablen aufweisen. Dem Studienprotokoll ist zusätzlich zu entnehmen, dass die Kategorisierung in weniger wichtig und sehr wichtig keinen Einfluss auf die Darstellung oder Behandlung der Confounder in der statistischen Analyse hat.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die grundsätzliche Vorgehensweise bei der Informationsbeschaffung des pU nach potenziellen Confoundern scheint nachvollziehbar und adäquat (vgl. Anhang A).

Aus der Literatursuche resultiert eine umfangreiche Liste potenzieller Confounder. Von den aus der Liste mittels Experteneinschätzung als sehr wichtig oder als weniger wichtig eingestuften Confoundern bzw. Operationalisierungen werden gemäß dem vom pU beschriebenen Vorgehen eine Reihe nicht für die AbD berücksichtigt. Das Vorgehen des pU ist nachvollziehbar, da er für jeden potenziellen Confounder auf Ebene der Operationalisierung eine Auswahl trifft. Der Ausschluss von Confoundern lediglich aufgrund der Experteneinschätzung als nicht wichtig, kann nicht nachvollzogen werden, da keine Begründung für diese Einschätzung vorliegt.

Der Confounder motorische Funktion soll über den höchsten motorischen Meilenstein, CHOP-INTEND und HFMSE operationalisiert werden. Insbesondere für den HFMSE bleibt entsprechend der Ausführungen in Abschnitt 2.2.3 unklar, wie mit Patientinnen und Patienten umgegangen wird, die das jeweilige Alter noch nicht erreicht haben, ab dem diese

Instrumente im SMARTCARE-Register erhoben werden. Dies ist vom pU in den Studienunterlagen zu spezifizieren.

Im Abgleich mit der 1. Version des Studienprotokolls [4] wurde die Liste potenziell relevanter Confounder um 3 Confounder erweitert: frühe Diagnosestellung, Mehrfacherkrankungen und körperliche Aktivität (operationalisiert durch Physiotherapie). Die frühe Diagnosestellung wird vom pU im Studienprotokoll als „Neugeborenencreening und frühe Diagnosestellung definiert“. Die Erfassung des Neugeborenencreenings (ja / nein) als Confounder erscheint nicht sinnvoll, da sich diese Variable innerhalb der vom G-BA gewählten Fragestellungen, in denen entweder alle Patientinnen und Patienten präsymptomatisch oder alle Patientinnen und Patienten symptomatisch sind, nicht unterscheiden kann. Wie der pU die frühe Diagnosestellung operationalisiert, ist unklar. Dies ist vom pU zu spezifizieren. Dass der pU Physiotherapie als Confounder erheben möchte, ist nachvollziehbar, da diese von den Klinikern als relevant (weniger wichtig) eingestuft wurde. Unklar bleibt allerdings, inwiefern die Physiotherapie den Confounder körperliche Aktivität abbildet. Dies ist klarzustellen.

Insgesamt ergeben sich in Abgleich mit der Confounderliste im Studienprotokoll zur AbD von Onasemnogen-Abepravovec und aufgrund der Tatsache, dass durch die aktualisierte systematische Confounder-Identifikation keine alleinig für Patientinnen und Patienten mit Typ 3 SMA potenziell relevanten Confounder identifiziert wurden, keine relevanten Lücken in der Liste potenzieller Confounder. Alle vom pU als relevant identifizierten Confounder werden bereits im SMARTCARE-Register erhoben. Für die oben genannten Confounder Physiotherapie, Mehrfacherkrankungen und frühe Diagnosestellung sind die konkreten Datenfelder im SMARTCARE-Register im Studienprotokoll (Tabelle 9) zu ergänzen bzw. zu spezifizieren.

Die Auflage des G-BA, die Studienunterlagen dahin gehend zu ergänzen, auf welche Weise die Confounder in das Modell zur Propensity Score Berechnung eingehen, wurde vom pU umgesetzt. Ebenfalls wurde vom pU angegeben, welchen Einfluss die Einteilung der Confounder in weniger wichtig und sehr wichtig für die Modellierung des Propensity Scores hat. Die Ausführungen sind sachgerecht.

2.2.15 Datenquelle: Meldezeitpunkte

Auflage des G-BA

Die Meldezeitpunkte in der gewählten Datenquelle müssen sicherstellen, dass die Daten der AbD für eine fristgerechte Einreichung der im Beschluss vom 21. Juli 2022 festgelegten Zwischenanalysen und des Dossiers für die erneute Nutzenbewertung zur Verfügung stehen. Diese Vorgabe muss in den Studienunterlagen hinterlegt werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In den Studienunterlagen wurde ergänzt, dass die Dateneingabe in den eCRF so zeitnah wie möglich erfolgen soll.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommene Änderung ist nicht ausreichend, um die Anforderung des G-BA umzusetzen. Es liegt im Verantwortungsbereich des pU sicherzustellen, dass zu den jeweils geplanten Datenschnitten für die Zwischenanalysen, alle bis dahin erhobenen Daten zur Verfügung stehen. Dies kann z. B. durch Studienmonitore unterstützt werden.

2.2.16 Datenquelle: Vollständigkeit der Daten

Auflage des G-BA

Die geplanten Plausibilitätsprüfungen zur Sicherstellung der Qualität und Vollständigkeit der Datenerhebung sind zu konkretisieren.

Angaben zu den Konsequenzen, die aus der geplanten Source Data Verification (SDV) gezogen werden, sind zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt, dass Patientendaten über präspezifizierte automatisierte sowie manuelle Prüfungen durch klinische Monitore validiert werden. Notwendige Anpassungen können dann direkt im eCRF durch das Prüfpersonal korrigiert werden. Darüber hinaus soll das Prüfpersonal besonders für die Vollständigkeit der Daten zu Endpunkten und Confoundern zu Baseline sensibilisiert werden. In den teilnehmenden Zentren werden regelmäßige Prüfungen zur Verbesserung der Datenqualität und -vollständigkeit durch wissenschaftliche Mitarbeiterinnen und Mitarbeiter durchgeführt. Auffälligkeiten in den Datensätzen werden vor jedem Statusbericht, sowie vor jeder Zwischenanalyse sowie der finalen Analyse in Data Review Meetings diskutiert.

Angaben zu Konsequenzen aus den Ergebnissen der SDV fehlen weiterhin.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Konkretisierungen in den Studienunterlagen zur Plausibilitätsprüfung sind weitgehend sachgerecht. Bei der Sensibilisierung für die Vollständigkeit der Daten zu Endpunkten und Confoundern sollten auch die Daten zu den Ein- und Ausschlusskriterien berücksichtigt werden. Angaben zu den Konsequenzen der SDV sind weiterhin zu ergänzen. Denkbar wären bei Auffälligkeiten beispielsweise spezifische Schulungen oder die Monitoringmaßnahmen in Zentren mit Auffälligkeiten zu erweitern.

2.2.17 Studiendesign: Indexdatum

Auflage des G-BA

Als Indexdatum, d. h. als Datum für den Beobachtungsstart, ist entsprechend des Vorgehens einer Target Trial Emulation der Tag der Therapieentscheidung (bzw. eine bestmögliche Näherung) festzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

In den Studienunterlagen wurde ergänzt, dass als Indexdatum der Tag der Therapieentscheidung festgelegt wird. In Fällen, in denen die Therapieentscheidung nicht dokumentiert ist, soll das Datum der erstmaligen Behandlung mit der Therapie, zu der der Patient zugeteilt wurde, als Indexdatum verwendet werden.

Die Visite, die am oder nächstgelegen vor dem Indexdatum oder am Indexdatum stattgefunden hat, soll als Baselinevisite herangezogen werden.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommene Änderung ist nicht sachgerecht. In den Fällen, in denen die Therapieentscheidung nicht dokumentiert ist und eine Brückentherapie eingesetzt wurde, stellt der Behandlungsbeginn mit dieser Brückentherapie das Indexdatum dar. Dies muss in den Studienunterlagen korrigiert werden. Darüber hinaus sind entsprechende Anstrengungen erforderlich, um den Anteil fehlender Angaben zum Tag der Therapieentscheidung zu minimieren. Die Baselinevisite soll möglichst zum Indexdatum stattfinden.

2.2.18 Studiendesign: Zuordnung zu den Behandlungsgruppen

Auflage des G-BA

Da eine Brückentherapie (z. B. bis zum Einsatz der Gentherapie) nicht auf Weiterbehandlung ausgelegt ist, ist festzuhalten, dass Patientinnen und Patienten, die eine Brückentherapie erhalten haben, der nachfolgenden Therapie zugeordnet werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU hat in den Studienunterlagen ergänzt, dass Patientinnen und Patienten, der Behandlungsgruppe entsprechend der nachfolgenden Therapie zugeordnet werden sollen, wenn die zuvor eingesetzte Therapie weniger als 3 Monate verabreicht wurde.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Der pU hat die Anforderungen des G-BA entsprechend umgesetzt. Zur Definition der Brückentherapie siehe Abschnitt 2.2.1.

2.2.19 Studiendesign: Fallzahlplanung

Auflage des G-BA

Die vorgelegte Fallzahlplanung ist nicht abschließend bewertbar.

Die Fallzahlplanungen für Patientinnen und Patienten mit SMA Typ 2 und mit Typ-3-SMA auf Grundlage des RULM sind anzupassen, da die Operationalisierung des RULM nicht sachgerecht ist (siehe Outcome, Morbidität).

Die Annahmen für die vorgelegte Fallzahlplanung für Patientinnen und Patienten mit Typ-1-SMA, sowie die Annahmen für die Effektstärke für Patientinnen und Patienten mit Typ-1- und Typ-2-SMA sind nicht nachvollziehbar. Die Annahmen sind genauer zu begründen und sofern erforderlich anzupassen.

Sofern sich für den initial gewählten Endpunkt trotz Überprüfung der zugrunde gelegten Annahmen unverhältnismäßig hohe Fallzahlen ergeben, sind seitens des pharmazeutischen Unternehmers weitere mögliche Fallzahlplanungen auf Basis anderer Nutzenendpunkte und ggf. auch auf Basis von Schadensendpunkten zu prüfen bzw. durchzuführen.

Aufgrund der hohen Unsicherheit bei der Abschätzung einer adäquaten Fallzahl im vorliegenden Verfahren ist in den Studienunterlagen zudem eine Überprüfung der Annahmen im Verlauf der Datenerhebung festzulegen.

Es ist zu ergänzen, auf welchen Verteilungsverhältnissen bezüglich der Behandlungsarme die jeweiligen Fallzahlplanungen beruhen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Für Fragestellung 3 und 4 (Patientinnen und Patienten mit Typ-2- und Typ-3-SMA) gibt der pU weiterhin an die Fallzahlplanung auf Grundlage des RULM vorzunehmen, operationalisiert als Änderung im Gesamtscore im Vergleich zu Baseline mit dem entsprechenden Effektmaß Cohen's d (als SMD).

Aufgrund unzureichender Daten zu den zu untersuchenden Therapien verzichtet der pU für beide Fragestellungen zunächst auf eine Fallzahlkalkulation und gibt lediglich Angaben zu den Standardannahmen, Signifikanzniveau ($\alpha = 2,5 \%$, 1-seitiger Test), Power (mindestens 80 %) sowie eine verschobene Nullhypothese von SMD = 0,5 an.

Für die Fragestellung 2 (Patientinnen und Patienten mit Typ-1-SMA) macht der pU Angaben zu den erwarteten Ereignisanteilen für Nusinersen und Onasemnogen-Abeparvovec, erläutert jedoch, dass unklar ist, wie die Aufteilung der Patientinnen und Patienten bezüglich beider Vergleichstherapien ist. Daher ist es aus Sicht des pU nicht möglich eine Annahme über den gesamten Vergleichsarm zu treffen. Somit verzichtet der pU in der vorliegenden Version auf

eine Fallzahlkalkulation und gibt lediglich Angaben zu den Standardannahmen, Signifikanzniveau ($\alpha = 2,5\%$, 1-seitiger Test) und die Power (mindestens 80 %), sowie eine verschobene Nullhypothese (Hazard Ratio [HR] = 0,5) an.

Der pU plant basierend auf den Ergebnissen der 1. Interimsanalyse Fallzahlkalkulationen für alle 4 Fragestellungen vorzunehmen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Da für alle 4 Fragestellungen aus Sicht des pU nicht genügend Evidenz für eine Fallzahlschätzung vorliegt, plant der pU im Rahmen der 1. Interimsanalyse Fallzahlschätzungen basierend auf den vorhandenen Ergebnissen. Dies betrifft auch Angaben zum Verteilungsverhältnis der Behandlungsmöglichkeiten. Das Vorgehen des pU ist nachvollziehbar, die diesbezügliche Anforderung des G-BA ist umgesetzt.

Unabhängig von der Operationalisierung (siehe Abschnitt 2.2.3) ist die vom pU gewählte verschobene Nullhypotesengrenze für den RULM in Fragestellung 3 und 4 (Patientinnen und Patienten mit Typ-2- und Typ-3-SMA) nicht sachgerecht. Richtwerte für angemessene Grenzen einer verschobenen Nullhypothese können die von Cohen [13] bzw. Rosenthal [14] angegebenen Grenzen für große bzw. sehr große Effekte sein ($SMD > 0,8$ bzw. $SMD > 1,3$).

2.2.20 Studiendesign: Abbruchkriterien

Auflage des G-BA

Vom pU ist zu ergänzen, wie für die Populationen vorgegangen werden soll, für die zum aktuellen Zeitpunkt noch keine Fallzahlplanung erfolgt ist.

Im Studienprotokoll ist zudem zu ergänzen, dass eine etwaige Entscheidung für einen Abbruch der AbD (sowie für eine Änderung der Fallzahlschätzung) in Abstimmung mit dem G-BA erfolgt.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU gibt an, dass im Rahmen der 1. Interimsanalyse basierend auf beobachteten Effektgrößen und Rekrutierungsanteilen eine Fallzahlschätzung erfolgen soll. Dies gilt, nach vorgenommenen Änderungen des pU im Abschnitt zur Fallzahlplanung (siehe Abschnitt 2.2.19), für alle betrachteten Patientenpopulationen. Laut pU wird die Studie abgebrochen, wenn die erwartete Power für den primären Endpunkt (und relevanter sekundäre Endpunkte) weniger als 60 % beträgt, hier bezogen auf die Ergebnisse der Interimsanalyse und den daraus resultierenden Fallzahlschätzungen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das vom pU beschriebene Vorgehen hinsichtlich eines Abbruchkriteriums bei Populationen für die zum aktuellen Zeitpunkt noch keine Fallzahlplanung erfolgt ist, ist zum Großteil sachgerecht, die Anforderung des G-BA ist erfüllt. Unklar bleibt bei den Ausführungen des pU, welche sekundären Endpunkte er als relevant ansieht, und wie er diese in die Entscheidung zum Abbruch einbeziehen will. Dabei ist darauf zu achten, dass die Entscheidung zum Abbruch in der Gesamtschau aller Daten erfolgt.

Angaben dazu, dass eine etwaige Entscheidung für einen Abbruch der AbD in Abstimmung mit dem G-BA erfolgt, finden sich nicht in den Studienunterlagen des pU. Diese Anforderung des G-BA ist nicht erfüllt und sollte ergänzt werden.

2.2.21 Studiendesign: Beobachtungsdauer

Auflage des G-BA

Für die AbD-Studie sind grundsätzlich alle Patientinnen und Patienten für mindestens 36 Monate nachzubeobachten, unabhängig von einem Behandlungswechsel.

Für den Endpunkt „ungeplante Hospitalisierungen“ ist entsprechend zu streichen, dass die Nachbeobachtung bei einem Behandlungswechsel mit dem Behandlungswechsel endet.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Im Studienprotokoll wurde ergänzt, dass alle Patientinnen und Patienten grundsätzlich 36 Monate nachbeobachtet werden sollen.

Der Endpunkt „ungeplante Hospitalisierungen“ wurde gestrichen. Für den primären Endpunkt bei präsymptomatischen Patientinnen und Patienten, der Anzahl UEs, die zu einer Hospitalisierung führen, ist in der primären Auswertungsstrategie entsprechend eines Treatment Policy Estimands, eine Berücksichtigung aller eingetretenen Ereignisse unabhängig vom Behandlungswechsel geplant. Lediglich in den ergänzenden Analysen zum Hypothetical Estimand werden für den Endpunkt UEs, die zu einer Hospitalisierung führen, nur Ereignisse vor Behandlungswechsel in die Analysen einbezogen.

Alle anderen UE-Endpunkte werden primär nach dem Hypothetical-Strategy-Ansatz, also einer verkürzten Beobachtungsdauer, ausgewertet. Eine zusätzliche Analyse nach dem Treatment-Policy-Ansatz wurde ergänzt, womit auch hier alle Ereignisse im Beobachtungszeitraum, unabhängig vom Behandlungswechsel, eingehen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen sind weitgehend sachgerecht. Für die Nutzenbewertung ist die Auswertung entsprechend des Treatment Policy Estimands die primär relevante Auswertungsstrategie.

2.2.22 Studiendesign: Angaben zum Verlauf der Datenerhebung

Auflage des G-BA

Für die geforderten Angaben zum Verlauf der Datenerhebung für den Zeitpunkt 6 Monate nach Studienbeginn sind, sofern möglich, auch bereits deskriptive Analysen von Daten zu prospektiv eingeschlossenen Patientinnen und Patienten vorzulegen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU hat im Studienprotokoll und im SAP ergänzt, dass zum ersten Statusreport 6 Monate nach Studienbeginn, wenn möglich, auch Analysen zu prospektiv eingeschlossenen Patientinnen und Patienten vorgelegt werden sollen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vorgenommene Änderung ist prinzipiell sachgerecht. Jedoch sollte, um dieses Ziel zu erreichen, auch der Zeitpunkt für den geplanten Datenschnitt angepasst werden. Dieser ist für diese Analyse weiterhin zu Studienbeginn geplant (und damit mit einer Vorlaufzeit von 6 Monaten). Da für die Verlaufserhebung lediglich deskriptive Analysen erforderlich sind, sollte eine kürzere Latenzzeit ausreichend sein, die ermöglicht, auch bereits zur ersten Verlaufserhebung Daten zu prospektiv eingeschlossenen Patientinnen und Patienten abzubilden.

2.2.23 Auswertung der Daten: Estimand

Auflage des G-BA

In den Studienunterlagen ist festzuhalten, dass der primäre Estimand der AbD-Studie der Treatment Policy Strategy entspricht und die Auswertung nach dem Intention-to-Treat(ITT)-Prinzip für alle patientenrelevanten Endpunkte beinhaltet. Bezüglich der Auswertung der unerwünschten Ereignisse wird zudem auf Ziffer 29 (Abschnitt 2.2.29) hingewiesen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die Studienunterlagen wurden dahin gehend geändert, dass die Estimands der primären Endpunkte der einzelnen Fragestellungen der Treatment Policy Strategy entsprechen. Weiterhin plant der pU, für Nebenwirkungsendpunkte zusätzlich den Estimand basierend auf der Treatment Policy Strategy zu verwenden (siehe Abschnitt 2.2.29). Bezüglich der sekundären Endpunkte macht der pU keine Angaben zum Estimand.

Zwar macht der pU bezüglich der Auswertungen nur teilweise Angaben zum Estimand, jedoch fehlt eine klare Beschreibung, ob die Auswertungen nach dem ITT-Prinzip erfolgen sollen. Der pU macht bezüglich der Auswertung der sekundären Endpunkte (RULM, HFMSE, CHOP-INTEND und 6-Minuten-Gehstrecke) operationalisiert als Änderung zu Baseline die Angabe, dass nur Patientinnen und Patienten in der Auswertung berücksichtigt werden, die einen Wert zu Baseline und zum auszuwertenden Zeitpunkt aufweisen. Für andere sekundäre Endpunkte, die die Zeit bis zum Erreichen eines interessierenden motorischen Meilensteines untersuchen, schränkt der pU die Analysepopulation auf Patientinnen und Patienten ein, die den jeweiligen motorischen Meilenstein zu Baseline nicht erreicht haben.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Anforderungen des G-BA sind vom pU nur teilweise umgesetzt. Grund hierfür ist, dass der pU nur für die primären Endpunkte als auch zusätzlich für die Nebenwirkungsendpunkte explizit benennt, dass er einen Estimand entsprechend der Treatment Policy Strategy verwenden will. Für sekundäre Endpunkte ist dies nicht umgesetzt. Dies ist vom pU in den Studienunterlagen zu ergänzen.

Die Auswertung von stetigen Endpunkten entsprechen nicht dem ITT-Prinzip, da nur Patientinnen und Patienten mit Baselinewert und mit beobachtetem Wert zum Auswertzeitpunkt berücksichtigt werden. Die Nichtberücksichtigung ist lediglich sachgerecht, wenn die fehlenden Werte aufgrund des nicht altersgerechten Einsatzes dieser Instrumente bedingt sind (jedoch unterliegt dieser Aspekt inhaltlichen Problemen, siehe Abschnitt 2.2.3). Eine Nichtberücksichtigung aus anderen Gründen führt zur Verletzung des ITT-Prinzips. Dies führt für die Fragestellungen 3 und 4 (Patientinnen und Patienten mit Typ-2-SMA bzw. Typ-3-SMA) dazu, dass sich die Angaben für den primären Endpunkt widersprechen, denn obwohl ein Estimand entsprechend der Treatment Policy Strategy geplant wird, ist das ITT-Prinzip ggf. verletzt. Dies ist nicht sachgerecht. Patientinnen und Patienten, die fehlende Werte aufweisen, obwohl das jeweilige Instrument für sie geeignet ist, müssen vom pU in den Analysen berücksichtigt werden.

Die Einschränkungen der Analysepopulationen für die Ereigniszeitanalysen methodisch sachgerecht, führen aber wie bei den stetigen Auswertungen zu inhaltlichen Problemen (siehe Abschnitt 2.2.3).

Zusätzlich sollten vom pU im SAP im Abschnitt zu den sekundären Endpunkten Angaben zum RULM ergänzt werden, da diese lediglich unter den primären Endpunkten aufgeführt sind.

2.2.24 Auswertung der Daten: Responderanalysen

Auflage des G-BA

Angaben zur geplanten Teststatistik für die geplanten Responderanalysen und dazu, welches Modell zur Berechnung verwendet werden soll, sind zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt die Teststatistiken für die geplanten Responderanalysen. Er gibt an, die Effektmaße absolute Risikoreduktion (ARR), Odds Ratio und relatives Risiko (RR) mit korrespondierenden p-Wert jeweils mittels einer logistischen Regression zu schätzen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU ist im Prinzip sachgerecht. Die Effektmaße ARR und RR werden jedoch nicht mittels einer logistischen Regression, also einer Regression mit der Logit-Funktion als Link-Funktion geschätzt. Dazu sind als Linkfunktion die Identität (ARR) bzw. die log-Funktion (RR) notwendig. Der pU sollte entsprechende Ergänzungen vornehmen.

2.2.25 Auswertung der Daten: stetige Auswertungen

Auflage des G-BA

Die fehlenden Angaben zum Effektmaß und zur Teststatistik für das geplante gemischte Modell für Messwiederholungen (Mixed Model Repeated measures [MMRM]) sind zu ergänzen.

Bezüglich der stetigen Auswertungen zum Endpunkt 6-Minuten-Gehtstrecke ist festzulegen, dass die Interpretation der Relevanz der Ergebnisse auf Basis der Skala des Instruments (d. h. vorliegend auf Basis der Wegstrecke) erfolgt.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt im SAP, dass zur Auswertung stetiger Endpunkte das Effektmaß Cohen's d mittels Effektschätzungen und Standardabweichungen für die Mittelwerte aus einem MMRM herangezogen wird.

Der pU gibt an, die Ergebnisse zur 6-Minuten-Gehstrecke sowohl für die einzelnen Therapiearme als auch für die Differenzen zwischen den Therapiearmen deskriptiv darzustellen. Zur Interpretation der Relevanz der Ergebnisse, d. h. welche Unterschiede in den Änderungen zum Studienbeginn klinisch relevant sind, äußert sich der pU nicht.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Eine genaue Definition des Effektmaßes Cohen's d im Zusammenhang mit der MMRM-Analyse ist weiterhin nicht beschrieben. Aus einem MMRM ergeben sich keine Standardabweichungen, sondern nur Standardfehler für die zu schätzenden Parameter des

MMRM. Es bleibt daher unklar, wie die gepoolte Standardabweichung für Cohen's d für beobachtete Werte angenähert wird. Des Weiteren fehlen immer noch Angaben zur Teststatistik, da die Parameterschätzungen aus einem MMRM verschiedene Effektschätzungen erlauben. Es können Unterschiede in den Änderungen zum Ausgangswert zwischen den Therapiearmen zu bestimmten Zeitpunkten geschätzt werden, aber auch Unterschiede in den über den Studienverlauf gemittelten Änderungen. Die Effektschätzungen können zu relevant unterschiedlichen Ergebnissen führen. Der pU hat die entsprechenden Angaben zu ergänzen.

Des Weiteren hat der pU die Interpretation der Relevanz der Ergebnisse auf Basis der Skala des Instruments 6-Minuten-Gehstrecke zu ergänzen.

2.2.26 Auswertung der Daten: Sensitivitätsanalysen

Auflage des G-BA

Die Planung von Heterogenitätsuntersuchungen hinsichtlich der Therapieoptionen im Vergleichsarm in der Datenauswertung als Sensitivitätsanalysen ist in den Studienunterlagen zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt im SAP getrennte Auswertungen für die Intervention gegenüber den Therapieoptionen im Kontrollarm. Diese werden für die primären Endpunkte in den 4 Populationen durchgeführt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU ist nicht sachgerecht. Es werden keine Heterogenitätsuntersuchungen beschrieben. Ein Ansatz wäre, die Daten im Sinne einer dreiarmligen Studie aufzufassen und entsprechende statistische Verfahren anzuwenden.

Weiterhin sollten die Sensitivitätsanalysen nicht nur für die primären Endpunkte, sondern auch für alle anderen patientenrelevanten Endpunkte durchgeführt werden.

2.2.27 Auswertung der Daten: Subgruppenanalysen

Auflage des G-BA

Die unterschiedliche Erfassung des geplanten Subgruppenmerkmals Alter („Alter zu Studieneinschluss“ und „Alter bei Diagnose“) ist für alle zu untersuchenden Populationen zu vereinheitlichen und für alle Populationen anzuwenden.

Eine inhaltliche Rationale für die Einteilung der Subgruppen auf Basis des Medians ist zu ergänzen.

Zudem sind Angaben zur geplanten Methodik für die Subgruppenanalysen und zum Umgang bzgl. der Interpretation der Ergebnisse zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Das Subgruppenmerkmal Alter wird entsprechend der überarbeiteten Studienunterlagen für alle Populationen als „Alter zu Behandlungsbeginn“ in 6 Kategorien (0 bis 18 Monate; 18 Monate bis 5 Jahre; 6 bis 11 Jahre; 12 bis 17 Jahre; 18 bis 25 Jahre und > 25 Jahre) untersucht.

Es wird weiterhin keine inhaltliche Rationale für die Einteilung der Subgruppen auf Basis des Medians angegeben (betrifft die Subgruppenmerkmale „CHOP-INTEND zu Baseline“ und „HFMSSE zu Baseline“). Im SAP wurde lediglich allgemein ergänzt, dass während der Ergebnisinterpretation berücksichtigt wird, ob es eine medizinische Rationale für eine Effektmodifikation gibt.

Der pU gibt an, entsprechende Interaktionstestungen mittels Likelihood-Ratio-Test durchführen zu wollen. Von den Analysen sollen Patientinnen und Patienten, die für das entsprechende Subgruppenmerkmal fehlende Werte aufweisen (fehlend bzw. unbekannt) ausgeschlossen werden. Eine Interpretation der p-Werte erfolgt nur, wenn in einer Subgruppe mehr als 10 Patientinnen und Patienten bzw. Ereignisse aufgetreten sind. Bei der Interpretation der Ergebnisse im Bericht werden sowohl die Konsistenz der Ergebnisse über die verschiedenen Endpunkte hinweg als auch die medizinische Rationale berücksichtigt.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die vom pU vorgenommenen Änderungen für das Merkmal Alter sind sachgerecht.

Nicht sachgerecht ist weiterhin die Einteilung einzelner Subgruppen auf Basis des Medians ohne inhaltliche Rationale. Es ist jeweils a priori ein inhaltlich begründeter Trennwert festzulegen, der nicht von den Ergebnissen der Studie abhängt. Andernfalls sind diese Subgruppenmerkmale verzichtbar.

Die Beschreibung der geplanten Methodik für die Subgruppenanalysen ist nicht vollständig. Zwar gibt der pU an, welche Teststatistik er hierfür verwenden will, jedoch fehlen Angaben zur konkreten Modellierung. Dies ist vom pU zu ergänzen. Weiterhin plant der pU Patientinnen und Patienten, die für das entsprechende Subgruppenmerkmal fehlende Werte aufweisen (fehlend bzw. unbekannt) aus der Analyse auszuschließen. Dieses Vorgehen soll um eine Sensitivitätsanalyse, welche diese Patientinnen und Patienten als Subgruppe mit einbezieht, erweitert werden.

Das beschriebene Vorgehen des pU zur Interpretation der Ergebnisse ist nachvollziehbar und sachgerecht, die Anforderung des G-BA wurde vom pU umgesetzt.

2.2.28 Auswertung der Daten: Propensity-Score-Verfahren

Auflage des G-BA

Beim Umgang mit extremen Gewichten ist das geplante Vorgehen durch geeignete Literatur zu begründen. Sofern keine geeignete Literatur vorliegt, ist das Verfahren anzupassen, z. B. durch das Trunkieren von extremen Gewichten.

Eine detaillierte Beschreibung bezüglich der Prüfung der Analysepopulation und Zielpopulation nach Propensity Score -Gewichtung ist vom pharmazeutischen Unternehmer zu ergänzen.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt im SAP, dass für die Auswertungen mittels Inverse Probability of Treatment Weighting (IPTW) und Fine Stratification Weights Patienten mit einem Propensity Score größer als 0,95 oder kleiner als 0,05 aus der Analyse ausgeschlossen werden. Eine Begründung oder eine Literaturreferenz für dieses Vorgehen wird nicht angegeben.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU ist nicht sachgerecht. Nach Stürmer 2021 [15] ist ein Vorgehen mit festen Schwellenwerten nicht empfehlenswert, da aufgrund der fixen Werte nicht ausreichend sicher ist, ob potenzielles Confounding ausreichend berücksichtigt wird. Stattdessen werden Verfahren mit relativen Schwellenwerten (bspw. unteres bzw. oberes 5 %-Perzentil) basierend auf den beobachteten Verteilungen der Propensity Scores bevorzugt [15,16].

2.2.29 Auswertung der Daten: unerwünschte Ereignisse (UE)

Auflage des G-BA

In den Studienunterlagen ist festzuhalten, dass für die UE-Endpunkte neben der geplanten Auswertung unter Verwendung einer Hypothetical Strategy auch Auswertungen unter Verwendung einer Treatment Policy Strategy durchgeführt werden.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt, dass die Auswertung gemäß der Treatment Policy Strategy (für den primären Endpunkt als Hauptanalyse, für die anderen Sicherheitsendpunkte als zusätzliche Analyse) durchgeführt wird.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das Vorgehen des pU ist sachgerecht.

2.2.30 Auswertung der Daten: Umgang mit fehlenden Werten

Auflage des G-BA

Die geplante Ersetzung des Monats für Patientinnen und Patienten mit Ereignis führt potenziell zu Verzerrungen und ist nicht sachgerecht. Diese Festlegung ist daher zu streichen. Vom pharmazeutischen Unternehmer ist stattdessen zu ergänzen, welche Anstrengungen unternommen werden, um die Rate fehlender Werte bei der Datumsangabe zu minimieren.

Es ist geplant, fehlende Werte nicht zu imputieren und eine Variable nicht als Confounder zu verwenden, wenn mehr als 40 % der Daten zu der jeweiligen Variablen fehlen. Seitens des pharmazeutischen Unternehmers sind die daraus resultierenden Konsequenzen für die Auswertungen und Interpretation der Daten zu diskutieren.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Die geplante Ersetzung des Monats für Patientinnen und Patienten mit Ereignis ist gestrichen. Zur Vermeidung fehlender Werte beschreibt der pU im Studienprotokoll die Verwendung von On-Site-Monitoring und Monitoringberichte zur Vermeidung fehlender Werte.

Zu den resultierenden Konsequenzen der Nichtberücksichtigung von Confoundern aufgrund zu hoher Anteile an fehlenden Werten für die Auswertung und Interpretation gibt der pU an, diese im Dossier zu beschreiben.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Die Streichung der geplanten Ersetzung des Monats für Patientinnen und Patienten mit Ereignis ist sachgerecht. Die beschriebenen Ansätze zur Vermeidung fehlender Werte ist sachgerecht.

Das Vorgehen bezüglich fehlender Confounder aufgrund zu hoher Anteile an fehlenden Werten ist nicht sachgerecht. Der pU gibt lediglich an, die Datenlage zu beschreiben. Mögliche Konsequenzen beim Ausschluss von Confoundern für die Interpretation der Ergebnisse werden jedoch nicht angeführt. In dieser Situation bleibt unklar, ob die Adjustierung ausreichend ist und damit die Anwendung eines Verfahrens mittels Propensity Score möglich ist. Sofern die Anwendung eines Propensity Score-basierenden Verfahrens nicht möglich ist, kann für die Nutzenbewertung ein naiver Vergleich ohne Adjustierung herangezogen werden. In diesem Fall sind die Konsequenzen bei der Ergebnisinterpretation zu beachten und zu beschreiben (Nachweis eines dramatischen Effekts) [17]. Entsprechende Anpassungen sind vom pU zu ergänzen.

2.2.31 Auswertung der Daten: Fehlende Angaben

Auflage des G-BA

Seitens des pharmazeutischen Unternehmers sind folgende fehlende Angaben entsprechend der Anforderungen des Forderungsbeschluss zu ergänzen:

- Umgang mit unplausiblen Daten und Ausreißern
- Angaben zur Überprüfung, inwieweit die parallel erhobenen, sowie zeitlich nicht parallel erhobenen Daten zu Nusinersen und Onasemnogen-Abeparvovec für eine gepoolte Analyse geeignet sind
- Angaben zur Überprüfung, inwieweit gegebenenfalls Daten zum Vergleich von Risdiplom gegenüber Nusinersen und Onasemnogen-Abeparvovec aus unterschiedlichen Datenquellen für eine gepoolte Analyse geeignet sind

Seitens des pharmazeutischen Unternehmers ist zur Vermeidung von Inkonsistenzen zu prüfen, ob der hier dargelegte Änderungsbedarf im Studienprotokoll zu entsprechenden Folgeänderungen im SAP führt und umgekehrt.

Umsetzung der Auflage des G-BA gemäß Darstellung des pU

Der pU beschreibt, dass unplausible Daten sowie Ausreißer in Data Review Meetings, die vor dem Statusbericht, der Zwischenanalyse und der finalen Analyse stattfinden, diskutiert und dokumentiert werden. Weiterhin erfolgt die Validierung der Patientendaten durch automatisierte, vorab festgelegte Prüfungen sowie durch klinische Monitore (siehe auch Abschnitt 2.2.16). Abfragelisten (Query-Listen) werden an den Prüfer gesendet, der die Daten direkt im eCRF korrigiert.

Zur Betrachtung von retrospektiv und prospektiv erhobenen Daten plant der pU Sensitivitätsanalysen durchzuführen.

Bewertung der vom pU vorgenommenen Änderungen durch das IQWiG

Das vom pU geplante Vorgehen bei unplausiblen Daten und Ausreißern ist bis auf den bereits in Abschnitt 2.2.16 beschriebenen Punkt zur Plausibilitätsprüfung sachgerecht.

Zwar gibt der pU an, die Eignung von gepoolten Analysen von retrospektiv und prospektiv erhobenen Patientinnen und Patienten mittels Sensitivitätsanalysen prüfen zu wollen, jedoch geht daraus nicht hervor, wie genau mit den erhaltenen Ergebnissen diese Eignung bewertet werden soll. Das Vorgehen des pU ist nicht vollständig beschrieben und somit nicht sachgerecht, entsprechende Spezifizierungen sind zu ergänzen.

Angaben zur Überprüfung, inwieweit gegebenenfalls Daten zum Vergleich von Risdiplom gegenüber Nusinersen und Onasemnogen-Abeparvovec aus unterschiedlichen Datenquellen

für eine gepoolte Analyse geeignet sind, führt der pU in seinen Studienunterlagen nicht auf. Da entsprechend der Angaben in den Studienunterlagen kein Einschluss unterschiedlicher Datenquellen geplant ist, sind entsprechende Betrachtungen nicht notwendig (siehe Abschnitt 2.2.13).

3 Literatur

1. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Risdiplam (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2022 [Zugriff: 12.09.2023]. URL: https://www.g-ba.de/downloads/40-268-8667/2022-07-21_AM-RL-XII_Risdiplam_AbD-004_Forderung_TrG.pdf.
2. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V; Risdiplam (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2022 [Zugriff: 12.09.2023]. URL: https://www.g-ba.de/downloads/39-261-5536/2022-07-21_AM-RL-XII_Risdiplam_AbD-004_Forderung_BAnz.pdf.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzept für eine anwendungsbegleitende Datenerhebung – Risdiplam; Rapid Report [online]. 2022 [Zugriff: 11.07.2023]. URL: https://www.iqwig.de/download/a21-131_anwendungsbegleitende-datenerhebung-risdiplam_rapid-report_v1-0.pdf.
4. F. Hoffmann-La Roche. Evaluation of a Real World Data Collection for the Reassessment of the Additional Benefit of Evrysdi (Risdiplam); Statistical Analysis Plan Version 1.0 [unveröffentlicht]. 2023.
5. Roche. Evaluation of a Real World Data Collection for the Reassessment of the Additional Benefit of Evrysdi (Risdiplam); NIS Protocol (Secondary Data Use); Version 1.0 [unveröffentlicht]. 2023.
6. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Anwendungsbegleitende Datenerhebung zu Risdiplam: Prüfung des Studienprotokolls (Version 1.0) und des statistischen Analyseplans (Version 1.0); 2. Addendum zum Projekt A21-131 [online]. 2023 [Zugriff: 27.05.2024]. URL: https://www.iqwig.de/download/a23-87_risdiplam_addendum-zum-projekt-a21-131_v1-1.pdf.
7. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Feststellung im Verfahren der anwendungsbegleitenden Datenerhebung und von Auswertungen nach § 35a Absatz 3b des Fünften Buches Sozialgesetzbuch (SGB V): Risdiplam (spinale Muskelatrophie) – Vorlage von Studienprotokoll und Statistischem Analyseplan [online]. 2024 [Zugriff: 27.05.2024]. URL: https://www.g-ba.de/downloads/39-261-6544/2024-04-04_AM-RL-XII_Risdiplam_AbD-004_Feststellung.pdf.

8. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Feststellung im Verfahren der anwendungsbegleitenden Datenerhebung und von Auswertungen nach § 35a Absatz 3b des Fünften Buches Sozialgesetzbuch (SGB V): Risdiplam (spinale Muskelatrophie) – Vorlage von Studienprotokoll und Statistischem Analyseplan [online]. 2024 [Zugriff: 27.05.2024]. URL: https://www.g-ba.de/downloads/40-268-10371/2024-04-04_AM-RL-XII_Risdiplam_AbD-004_Feststellung_TrG.pdf.
9. F. Hoffmann-La Roche. Evaluation of a Real World Data Collection for the Reassessment of the Additional Benefit of Evrysdi (Risdiplam); Statistical Analysis Plan Version 2.0 [unveröffentlicht]. 2024.
10. Roche. Evaluation of a Real World Data Collection for the Reassessment of the Additional Benefit of Evrysdi (Risdiplam); NIS Protocol (Secondary Data Use); Version 2.0 [unveröffentlicht]. 2024.
11. Proud CM, Mercuri E, Finkel RS et al. Combination disease-modifying treatment in spinal muscular atrophy: A proposed classification. *Ann Clin Transl Neurol* 2023; 10(11): 2155-2160. <https://doi.org/10.1002/acn3.51889>.
12. Bayley N. Bayley Scales of Infant and Toddler Development; Administration Manual. San Antonio: PsychCorp, Harcourt Assessment; 2006.
13. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum; 1988.
14. Rosenthal JA. Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 1996; 21(4): 37-59. https://doi.org/10.1300/J079v21n04_02.
15. Stürmer T, Webster-Clark M, Lund JL et al. Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study. *Am J Epidemiol* 2021; 190(8): 1659-1670. <https://doi.org/10.1093/aje/kwab041>.
16. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015; 34(28): 3661-3679. <https://doi.org/10.1002/sim.6607>.
17. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 21.09.2023]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf.

18. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Review des Studienprotokolls und statistischen Analyseplans zur anwendungsbegleitenden Datenerhebung von Onasemnogen-Abeprarvovec - Addendum zum Auftrag A20-61 [online]. 2021 [Zugriff: 09.06.2024]. URL: https://www.iqwig.de/download/a21-107_anwendungsbegleitende-datenerhebung-onasemnogen-abeparvovec-pruefung-protokoll-und-sap_addendum-zum-auftrag-a20-61_v1-0.pdf.

Anhang A Prüfung der Informationsbeschaffung zur Identifizierung von Confoundern

Informationsbeschaffung

Zur Identifikation der relevanten Confounder legt der pU einen Bericht über eine systematische Literatursuche zur Identifikation und Validierung von Confoundern für das gesamte relevante Anwendungsgebiet der vorliegenden AbD vor. Der pU führt eine systematische Literaturrecherche in MEDLINE, Cochrane Database of Systematic Reviews sowie ergänzend eine Handsuche in verschiedenen Datenbanken und spezifischen Websites deutscher und internationaler Fachgesellschaften nach Leitlinien, Empfehlungen, Systematischen Übersichten (SÜs) und Metaanalysen durch.

Relevante Leitlinien und Empfehlungen sollen Informationen zu prognostischen Faktoren liefern, während die in SÜs und Metaanalysen enthaltenen Studien mindestens einen patientenrelevanten Endpunkt berichten müssen. Die in den relevanten Dokumenten eingeschlossenen Studien wurden extrahiert und potenzielle Confounder, Risikofaktoren sowie prognostischen Faktoren erfasst. Die so identifizierten potenziellen Confounder wurden durch Interviews mit 7 klinischen Expertinnen und Experten validiert.

Kommentar zur Informationsbeschaffung

Die dargestellte Informationsbeschaffung zur Identifikation von Confoundern wurde bereits im Verfahren zur AbD von Onasemnogen-Abeparvovec als geeignet beurteilt [18].

Analog zum Verfahren zur AbD von Onasemnogen-Abeparvovec führt der pU eine Suche nach Leitlinien und systematischen Übersichten / Metaanalysen in bibliografischen Datenbanken durch. Obwohl in der Methodik beschrieben wurde, dass die Confounder aus den eingeschlossenen Primärstudien extrahiert wurden, kann dies bei der Ergebnisdarstellung nicht nachvollzogen werden. Diese beinhaltet ausschließlich Listen relevanter systematischer Übersichten, Metaanalysen, Leitlinien und Empfehlungen. Aufgrund der vorliegenden Informationen kann die Informationsbeschaffung nicht vollständig bewertet werden.

Die grundsätzliche Vorgehensweise bei der Informationsbeschaffung des pU nach Confoundern scheint jedoch nachvollziehbar und adäquat. Es kann daher davon ausgegangen werden, dass eine ausreichend vollständige Liste an potenziell relevanten Confoundern identifiziert wurde.